# Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution+ (GC+) Property Models and Uncertainty Analysis


| | |
|---|---|
| Journal: | *Journal of Chemical Information and Modeling* |
| Manuscript ID: | Draft |
| Manuscript Type: | Article |
| Date Submitted by the Author: | n/a |



| | |
|---|---|
| Complete List of Authors: | Hukkerikar, Amol; Technical University of Denmark, Chemical and Biochemical Engineering<br>Kalakul, Sawitree; The Petroleum and Petrochemical College,<br>Sarup, Bent; Alfa Laval Copenhagen A/S, Vegetable Oil Technology Business Unit<br>Young, Douglas; U.S. Environmental Protection Agency, Clean Processes Branch, National Risk Management Research Laboratory<br>Sin, Gurkan; Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering<br>Gani, Rafiqul; Technical University of Denmark, Department of Chemical Engineering |

# Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution$^{+}$ ($GC^{+}$) Property Models and Uncertainty Analysis

*Amol Shivajirao Hukkerikar* [a], *Sawitree Kalakul* [b], *Bent Sarup* [c], *Douglas M. Young* [d], *Gürkan Sin* [a], *and Rafiqul Gani* [a,*]

[a] Computer Aided Process-Product Engineering Center (CAPEC), Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark.

[b] The Petroleum and Petrochemical College, Chulalongkorn University, Bangkok 10330, Thailand.

[c] Vegetable Oil Technology Business Unit, Alfa Laval Copenhagen A/S, Maskinvej 5, DK-2860, Søborg, Denmark.

[d] U.S. EPA, National Risk Management Research Laboratory, Cincinnati, Ohio 45268.

[*]Corresponding author: E-mail address: rag@kt.dtu.dk (R. Gani); Phone no: +45-45252882; Fax no: +45-45932906.

**ABSTRACT:** The aim of this work is to develop group-contribution$^+$ ($GC^+$) method (combined group-contribution (GC) method and atom connectivity index (CI) method) based property models to provide reliable estimations of environment-related properties of organic chemicals together with uncertainties of estimated property values. For this purpose, a systematic methodology for property modeling and uncertainty analysis is used. The methodology includes a parameter estimation step to determine parameters of property models and an uncertainty analysis step to establish statistical information about the quality of parameter estimation, such as the parameter covariance, the standard errors in predicted properties, and the confidence intervals. For parameter estimation, large data-sets of experimentally measured property values of wide range of chemicals (hydrocarbons, oxygenated chemicals, nitrogenated chemicals, poly-functional chemicals, etc.) taken from the database of US Environmental Protection Agency (EPA) and from the database of USEtox$^{TM}$ is used. For property modeling and uncertainty analysis, the Marrero and Gani GC method and atom connectivity index method have been considered. In total 22 environment-related properties, which include the fathead minnow 96-hr $LC_{50}$, *daphnia magna* 48-hr $LC_{50}$, oral rat $LD_{50}$, aqueous solubility, bioconcentration factor, permissible exposure limit (OSHA-TWA), photochemical oxidation potential, global warming potential, ozone depletion potential, acidification potential, emission to urban air (carcinogenic and non-carcinogenic), emission to continental rural air (carcinogenic and non-carcinogenic), emission to continental fresh water (carcinogenic and non-carcinogenic), emission to continental sea water (carcinogenic and non-carcinogenic), emission to continental natural soil (carcinogenic and non-carcinogenic), emission to continental agricultural soil (carcinogenic and non-carcinogenic) have been modeled and analysed. The application of the developed property models for the estimation of environment-related properties and uncertainties of the estimated property values is highlighted through an illustrative example. The developed property models provide reliable estimates of

environment-related properties needed to perform process synthesis, design, and analysis of sustainable chemical processes and allow one to evaluate the effect of uncertainties of estimated property values on the calculated performance of processes giving useful insights into quality and reliability of the design of sustainable processes.

**KEYWORDS:** group-contribution$^+$ ($GC^+$) method, uncertainty analysis, potential environmental impact, life cycle impact assessment, sustainable process design.

## INTRODUCTION

Currently, there is a great deal of interest in the development of computer aided methods and tools for the process synthesis, design, and analysis of sustainable processes. The design of sustainable processes require the satisfying of various conditions (or constraints) such as, increased productivity, minimum energy consumption, reduction in raw materials, recovery of products, and minimum generation of pollution.[1] This task can be effectively accomplished by using a chemical process simulator (to perform mass and energy balances for the concerned process) together with the waste reduction (WAR) algorithm[2-3] to obtain a quantitative measure of the potential environmental impact (PEI) which, as part of the life cycle assessment (LCA) of process synthesis and design, contributes to identifying sustainable processing paths and design alternatives. The PEI is a relative measure of the potential for a chemical to have an adverse effect on human health and the environment. Several studies in literature[4-6] have reported the application of the WAR algorithm for generating sustainable process design alternatives and deciding on sustainable process designs that are environmentally friendly and economically attractive. In the WAR algorithm, the total PEI of a process is evaluated based on the following eight categories of potential impacts: (i) human toxicity potential by ingestion, calculated using oral rat $LD_{50}$; (ii) human toxicity potential by exposure both dermal and inhalation, calculated using

permissible exposure limit (OSHA-TWA); (iii) terrestrial toxicity potential, calculated using oral rat $LD_{50}$; (iv) aquatic toxicity potential, calculated using fathead minnow 96-hr $LC_{50}$; (v) global warming potential; (vi) ozone depletion potential; (vii) photochemical oxidation potential; and (viii) acidification potential. Therefore, the basis for the quantification of PEI is a set of environment-related properties (such as fathead minnow 96-hr $LC_{50}$, oral rat $LD_{50}$, global warming potential etc.) of chemical substances involved in the process. The USEtox™ model is an environment model for characterisation of human and ecotoxicological impacts in life cycle impact assessment (LCIA) and Comparative Risk Assessment (CRA) and is designed to describe the fate, exposure and effects of chemicals.[7-8] The USEtox™ model calculates characterisation factors for carcinogenic impacts, non-carcinogenic impacts, and total impacts (carcinogenic + non-carcinogenic) based on the chemical emissions to urban air, rural air, freshwater, sea water, agricultural soil and/or natural soil. For many chemicals of interest the experimental data of environment-related properties is not available since the measurement of these properties are extremely time consuming and expensive. Also, processes that deal with the synthesis of new chemicals require a suitable property prediction method in order to obtain reliable estimates of environment-related properties of new chemicals. A review article by Boethling et al.[9] discusses available experimental data sources and various estimation methods including group-contribution (GC) methods, methods based on quantitative structure-property relationships (QSPR), and correlation equations, to name a few for obtaining values of environment-related properties of chemicals.

For the estimation of properties of organic chemicals, GC methods such as those reported by Joback and Reid[10], Lydersen[11], Constantinou and Gani[12], and Marrero and Gani[13], have been widely employed to obtain the needed property values since these methods provide the advantage of quick estimates without requiring substantial computational work. In GC methods, the property of a chemical is a function of structurally dependent parameters, which

are determined as a function of the frequency of the groups representing the chemical and their contributions. Among GC methods for estimation of properties of chemicals, the Marrero and Gani (MG) method[13] is well-known. The MG method allows estimation of properties based exclusively on the molecular structure of the chemical and exhibits a good accuracy and a wide range of applicability covering chemical, biochemical, and environment-related chemicals. Note, that for reliable estimation of properties of chemicals using a GC method, the user needs: (i) a property model; (ii) group definitions (model parameters of the selected property model) and their contributions; and (iii) a tool to quantify uncertainties (prediction errors) of estimated property values in order to check the quality (reliability) of estimation. In many cases, however, the selected property model may not have all the needed model parameters (that is, groups describing the molecular structure of a given chemical) and/or their contributions. In such cases where the molecular structure of a given chemical is not completely described by any of the available groups, the atom connectivity index (CI) method has been employed together with MG method to create the missing groups and to predict their contributions.[14] This combined approach has led to the development of group contribution$^+$ ($GC^+$) method of a wider application range than before since the missing groups and their contributions can now be easily predicted through the regressed contributions of connectivity indices.[14]

There are numerous LCA software tools available (for example, SimaPro, GaBi etc.) for quantification of potential impact that the processes would have on the environment on average. Most of these tools have built-in databases containing properties of chemicals needed for the environmental-impact analysis. However, for chemicals that are not included in the database, a suitable property prediction method is necessary to obtain the needed environment-related property values which will allow one to perform synthesis, design, and analysis of sustainable chemical processes. For the estimation of fathead minnow 96-hr $LC_{50}$

and aqueous solubility, various GC-methods have been developed. Martin and Young[15] developed a GC method to correlate the acute toxicity (96-hr $LC_{50}$) to the fathead minnow using 397 organic chemicals based on the multi-linear regression and computational neural networks approach for the parameter estimation. Casalegno et al.[16] used a diatomic fragment approach based GC method to correlate the acute toxicity (96-hr $LC_{50}$) of 607 organic chemicals. For the estimation of aqueous solubility, Marrero and Gani[17] developed a GC method using a three-level parameter estimation approach (with a dataset of 2087 organic chemicals used for the regression purpose) and this method requires only molecular structural information for the estimation of aqueous solubility. There are several other GC methods available for estimation of aqueous solubility (Klopman and Zhu[18]; Kühne et al.[19]). For the estimation of oral rat $LD_{50}$ and bio-concentration factor (BCF), the more common approach has been to employ correlation equations (for example, bio-concentration factor for a chemical is estimated using known value of its octanol/water partition coefficient). Moreover, Martin et al.[20] have developed a hierarchical clustering technique to predict a variety of endpoints, including oral rat $LD_{50}$, BCF, aqueous solubility, fathead minnow $LC_{50}$ that combines group contributions with descriptors from graph theory. Software platforms have been developed both in the U.S. (US EPA 2012[21-22]) and in Europe (Isitutio Mario Negri 2012[23]) to predict these same endpoints. The application range and capability of these estimation equations is limited by the availability of the required property values. To the best of our knowledge, there are no GC methods reported in the literature for the estimation of following environment-related properties: permissible exposure limit (OSHA-TWA), global warming potential, photochemical oxidation potential, ozone depletion potential, acidification potential, emission to urban air (carcinogenic and non-carcinogenic), emission to continental rural air (carcinogenic and non-carcinogenic), emission to continental fresh water (carcinogenic and non-carcinogenic), emission to continental sea water (carcinogenic and

non-carcinogenic), emission to continental natural soil (carcinogenic and non-carcinogenic), and emission to continental agricultural soil (carcinogenic and non-carcinogenic). In addition to the accurate estimation of environment-related properties, it is also important to know the uncertainties (for example, prediction error in terms of 95% confidence interval) of the estimated property values that arise due to uncertainties of the regressed parameters of the selected property model. With this information, it is possible to evaluate the effect of these uncertainties on the calculated potential impact that the processes would have on the environment and to verify the quality and reliability of the design of sustainable processes.

Motivated by the preceding literature review and by the need of reliable estimation of environment-related properties in synthesis, design, and analysis of sustainable processes, this work aims to develop property prediction models based on the $GC^+$ approach (combined GC method and CI method) to provide reliable estimates of environment-related properties together with uncertainties of the estimated property values. For this purpose, a systematic methodology for property modeling and uncertainty analysis developed by Hukkerikar et al.[24] is used. The methodology includes a parameter estimation step to determine parameters (group/atom contributions, adjustable parameters, and a universal parameter) of property models and an uncertainty analysis step to establish statistical information about the quality of parameter estimation, such as the parameter covariance, the standard errors in predicted properties, and the confidence intervals. For property modeling with a GC method, the MG method[13] has been considered. For property modeling with a CI method, the models proposed by Gani et al.[14] have been considered. For parameter estimation, large data-sets of experimentally measured property values of wide range of chemicals taken from the database of US Environmental Protection Agency (EPA) and from the database of USEtox™ is used. In total 22 environment-related properties, which include the fathead minnow 96-hr $LC_{50}$ ($LC_{50}$(FM)), *daphnia magna* 48-hr LC50 ($LC_{50}$(DM)), oral rat $LD_{50}$, aqueous solubility

(Log$W_s$), bioconcentration factor (BCF), permissible exposure limit (PEL(OSHA-TWA)), photochemical oxidation potential (PCO), global warming potential (GWP), ozone depletion potential (ODP), and acidification potential (AP), emission to urban air (carcinogenic ($EUA_C$) and non-carcinogenic ($EUA_{NC}$)), emission to continental rural air (carcinogenic ($ERA_C$) and non-carcinogenic ($ERA_{NC}$)), emission to continental fresh water (carcinogenic ($EFW_C$) and non-carcinogenic ($EFW_{NC}$)), emission to continental sea water (carcinogenic ($ESW_C$) and non-carcinogenic ($ESW_{NC}$)), emission to continental natural soil (carcinogenic ($ENS_C$) and non-carcinogenic ($ENS_{NC}$)), emission to continental agricultural soil (carcinogenic ($EAS_C$) and non-carcinogenic ($EAS_{NC}$)) have been modeled and analysed.

The paper first gives a brief overview of the methodology for property modeling and uncertainty analysis; followed by model performance statistics; and finally, application of the developed property models for estimation of environment-related properties. Tables containing list of property model parameters together with parameter values, due to their large size, are provided as supporting information.

# METHODS AND TOOLS FOR PROPERTY MODELING AND UNCERTAINTY ANALYSIS

**MG Group-Contribution Method.** In the MG method[13] the property estimation is performed at three levels. The first level has a large set of simple groups that allow for the representation of a wide variety of organic chemicals. However, these groups only partially capture the proximity effects and are unable to distinguish among isomers. The second level of estimation involves groups that provide better description of proximity effects and can differentiate among isomers. Hence, second level of estimation is intended to deal with polyfunctional, polar or non-polar, and cyclic chemicals. The third level estimation includes groups that provide more structural information about molecular fragments of chemicals whose description is insufficient through the first- and second-order groups; hence, this level allows estimation of complex heterocyclic and polyfunctional acyclic chemicals. The MG method includes 220 first-order groups, 130 second-order groups, and 74 third-order groups to represent the molecular structure of the organic chemicals. The property prediction model to estimate the properties of organic chemicals employing MG method has the form[13],

$$f(X) = \sum_i N_i C_i + w \sum_j M_j D_j + z \sum_k E_k O_k \qquad (1)$$

The function $f(X)$ is a function of property $X$ and it may contain additional adjustable model parameters (universal constants) depending on the property involved. In eq. (1), $C_i$ is the contribution of the first-order group of type-$i$ that occurs $N_i$ times. $D_j$ is the contribution of the second-order group of type-$j$ that occurs $M_j$ times. $E_k$ is the contribution of the third-order group of type-$k$ that has $O_k$ occurrences in a component. Eq. (1) is a general model for all the properties and the definition of $f(X)$ is specific for each property $X$ (see Table- 2 for definition of $f(X)$ for models of environment-related properties). For determination of the contributions, $C_i$, $D_j$, and $E_k$, Marrero and Gani[13] suggested a three-step regression procedure.

- *Step 1*: In this step, the constants *w* and *z* are assigned zero values because only contributions of the first-order groups are estimated, that is, the first-order groups, $C_i$ and the additional adjustable parameters of the model.

$$f(X) = \sum_i N_i C_i \tag{2}$$

- *Step 2*: In this step, the constants *w* and *z* are assigned unity and zero values, respectively, because only first and second-order groups are considered. The regression is performed (by keeping fixed the $C_i$ and the adjustable parameters obtained from step 1) to determine the contributions of the second-order groups, $D_j$.

$$f(X) = \sum_i N_i C_i + w\sum_j M_j D_j \tag{3}$$

- *Step 3*: In this step, both *w* and *z* are set to unity and regression is performed (by keeping fixed the obtained $C_i$, $D_j$, and the adjustable parameters obtained from steps 1 and 2) to determine the contributions of the third-order groups, $E_k$ (see eq. 1).

In this way, the contributions of higher levels act as corrections to the approximations of the lower levels. Hukkerikar et al.[24] discussed a new approach for estimating the contributions, $C_i$, $D_j$, and $E_k$ based on the simultaneous regression method in which regression is performed by considering all the terms of the eq. (1) to obtain contributions of first-, second-, and third-order groups in a single regression step.

**Atom Connectivity Index (CI) Method.** This method employs the following model for the estimation of properties of organic chemicals[14]:

$$f(X) = \sum_i a_i A_i + b\left({}^{v}\chi^{0}\right) + 2c\left({}^{v}\chi^{1}\right) + d \tag{4}$$

Where $a_i$ is the contribution of the atom of type-*i* that occurs $A_i$ times in the molecular structure, ${}^{v}\chi^{0}$ is the zeroth-order (atom) valence connectivity index, ${}^{v}\chi^{1}$ is the first-order

(bond) valence connectivity index, *b* and *c* are adjustable parameters, and *d* is a universal parameter. Please note that *f(X)* of models in the MG method[13] and in the CI method[14], (i.e. left hand side of eq. (1) and eq. (4)) has the same functional form for a particular pure component property *X* and the values of universal constants for the CI models are the same as those for the GC models.

**Database.** For the estimation of property model parameters, large experimental data-sets of organic chemicals of various classes (hydrocarbons, oxygenated components, nitrogenated components, poly-functional components, etc.) from the database of US Environmental Protection Agency (EPA) and from the database of USEtox$^{TM}$ is used. The details of data-set of each property in terms of number of organic chemicals belonging to various classes are given in Table 1(a) (data-sets from US Environmental Protection Agency (EPA)) and in Table 1(b) (data-sets from USEtox$^{TM}$).

**Table 1(a). Description of the US EPA Data-sets Used for the Regression Purpose**

| class of chemicals | $LC_{50}$ (FM) | $LC_{50}$ (DM) | $LD_{50}$ | $LogW_s$ | BCF | PCO | PEL | GWP | ODP | AP |
|---|---|---|---|---|---|---|---|---|---|---|
| hydrocarbons | 32 | 19 | 69 | 236 | 79 | 337 | 98 | 0 | 0 | 0 |
| oxygenated | 238 | 54 | 1382 | 1110 | 76 | 244 | 127 | 1 | 0 | 0 |
| nitogenated | 80 | 24 | 397 | 244 | 57 | 8 | 45 | 0 | 0 | 0 |
| chlorinated | 48 | 37 | 111 | 274 | 77 | 23 | 41 | 5 | 3 | 5 |
| fluorinated | 1 | 0 | 3 | 21 | 1 | 5 | 4 | 23 | 0 | 0 |
| brominated | 10 | 4 | 14 | 47 | 15 | 5 | 7 | 2 | 1 | 0 |
| iodinated | 1 | 0 | 5 | 17 | 0 | 0 | 1 | 0 | 0 | 0 |
| phosphorous containing | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| sulfonated | 9 | 8 | 24 | 19 | 5 | 0 | 15 | 0 | 0 | 0 |
| silicon containing | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| multifunctional | 390 | 174 | 3984 | 2711 | 352 | 17 | 87 | 20 | 24 | 5 |
| total number of chemicals | 809 | 320 | 5995 | 4681 | 662 | 639 | 425 | 51 | 28 | 10 |

**Table 1(b). Description of the USEtox™ Data-sets Used for the Regression Purpose**

| class of chemicals | $EUA_C$ | $EUA_{NC}$ | $ERA_C$ | $ERA_{NC}$ | $EFW_C$ | $EFW_{NC}$ | $ESW_C$ | $ESW_{NC}$ | $ENS_C$ | $ENS_{NC}$ | $EAS_C$ | $EAS_{NC}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| hydrocarbons | 25 | 14 | 18 | 16 | 19 | 14 | 19 | 16 | 18 | 16 | 20 | 16 |
| oxygenated | 107 | 56 | 96 | 60 | 98 | 57 | 101 | 60 | 96 | 58 | 97 | 58 |
| nitogenated | 29 | 14 | 27 | 14 | 27 | 13 | 26 | 15 | 27 | 15 | 27 | 14 |
| chlorinated | 46 | 23 | 43 | 26 | 44 | 27 | 45 | 32 | 45 | 30 | 43 | 28 |
| fluorinated | 4 | 1 | 4 | 1 | 4 | 1 | 4 | 1 | 4 | 1 | 4 | 1 |
| brominated | 6 | 2 | 5 | 2 | 4 | 2 | 5 | 2 | 5 | 3 | 5 | 3 |
| iodinated | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| phosphorous containing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| sulfonated | 3 | 1 | 3 | 1 | 3 | 1 | 3 | 1 | 3 | 1 | 3 | 1 |
| silicon containing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| multifunctional | 236 | 230 | 274 | 229 | 273 | 230 | 274 | 233 | 262 | 238 | 271 | 231 |
| total number of chemicals | 456 | 341 | 470 | 349 | 472 | 345 | 477 | 360 | 460 | 362 | 470 | 352 |

**Parameter Estimation and Uncertainty Analysis (Maximum-Likelihood Estimation)**. The following discussion on parameter estimation and uncertainty analysis is based on the methodology discussed by Hukkerikar et al.[24] Let the property prediction model be represented by $f$ and the model parameters (group/atom contributions, adjustable parameters, and universal parameter) by $\boldsymbol{P}$. The minimization of a cost function, $S(\boldsymbol{P})$, defined as the sum of the squares of the difference between the experimental value, $X^{exp}$, and evaluated property value, $X^{pred}$, provides the values of unknown parameters $\boldsymbol{P}^*$. This implies that $\boldsymbol{P}^*$ is a set of model parameter values obtained at the minimum value of the cost function value.

$$S(\boldsymbol{P}) = min \sum_{j=1}^{N} \left( X_j^{exp} - X_j^{pred} \right)^2 \quad (5)$$

The subscript $j$ indicates the chemical evaluated and $N$ is the total number of chemicals included in the evaluation. After the estimation of the model parameters, uncertainty analysis is performed to quantify the model prediction errors. In this work, since the proposed models

for environment-related properties are linear in nature, the following discussion is intended to provide information on linear least squares theory. For linear least squares, the covariance matrix of the estimated model parameters, $COV(P^*)$, is given by (Seber and Wild[25]),

$$COV\left(\boldsymbol{P}^*\right) = \frac{SSE}{\nu}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1} \quad (6)$$

Where, *SSE* is the sum of squared errors obtained from the least-squares parameter estimation method, $v$ is the degrees of freedom (that is, the total number of measurements, $n$ minus the number of unknown parameters, $m$). For the GC model with linear form of $f(X)$, $\boldsymbol{A}$ is the matrix containing frequencies (or occurrences) of groups used to represent the chemicals in the data-set used for the regression. For the CI model with linear form of $f(X)$, $\boldsymbol{A}$ is the matrix containing frequencies of atoms and zeroth-order and first-order connectivity index for each chemical included in the data-set. The covariance matrix computed using eq. (6) is used for assessing the quality of the parameter estimation. The diagonal elements of this matrix are the variances of the errors of the parameter estimates and the off-diagonal elements are the covariances between the parameter estimation errors.

The confidence interval of the parameters, $\boldsymbol{P}^*$, at $\alpha_t$ significance level is given as (Seber and Wild[25]; Sin et al.[26]),

$$\boldsymbol{P}^*_{1-\alpha_t} = \boldsymbol{P}^* \pm \sqrt{\mathrm{diag}\left(COV\left(\boldsymbol{P}^*\right)\right)} \;.\; t\left(\nu,\ \alpha_t/2\right) \quad (7)$$

In eq. (7), $t(v,\ \alpha_t\ /2)$ is the t-distribution value corresponding to the $\alpha_t/2$ percentile ($\alpha_t$ is usually a value of 0.05) and $\mathrm{diag}(COV(\boldsymbol{P}^*))$ represents the diagonal elements of $COV(\boldsymbol{P}^*)$. The t-distribution value is obtained from the probability distribution function of Students t-distribution, (Abramowitz and Stegun[27]), $P_r(t, v)$, and is given as,

$$0 = \sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)^{-1} \int_{-t}^{t} \left(1 + x^2 / \nu\right)^{-\frac{1}{2}(\nu+1)} dx \; - \; P_r(t, \nu) \tag{8}$$

Where $x = \dfrac{\nu}{\nu + t^2}$ and $B$ (1/2, ν/2) is the beta function. For 95% confidence interval calculation, the value of $P_r(t, \nu)$ is 0.95. The t-distribution value can also be obtained using the "tinv" function available in MatLab.

The confidence interval of the predicted property value, $X^{pred}$, at $\alpha_t$ significance level is given as,

$$X^{pred}_{1-\alpha_t} = X^{pred} \pm \sqrt{\text{diag}\left( J\left(\boldsymbol{P}^*\right) COV\left(\boldsymbol{P}^*\right) J\left(\boldsymbol{P}^*\right)^T \right)} \; . \; t\left(\nu, \alpha_t / 2\right) \tag{9}$$

Where, the Jacobian matrix $J(\boldsymbol{P}^*)$ calculated using $\partial f / \partial \boldsymbol{P}^*$ represents the local sensitivity of the property model $f$ to variations in the estimated parameter values $\boldsymbol{P}^*$. The $X^{pred}_{1-\alpha_t}$ calculated from eq. (9) can be used to assess the reliability of the prediction (when experimental data is available for the property). If the experimental value of the property is within the calculated confidence interval, then the property prediction method is verified as reliable. When experimental data is unavailable, the calculated confidence interval provides a measure of the likely prediction error (uncertainty) of the predicted property value. This information can be used in the design and analysis of sustainable processes to take into account the effect of uncertainties of predicted property values on the calculated impact that the processes would have on the environment (and hence on the decision of selection of sustainable process design).

**Statistical Performance Indicators.** The statistical significance of the developed correlations in this work is based on the following performance indicators (Hukkerikar et al.[24]).

- *Standard deviation* (SD): This parameter measures the spread of the data about its mean value μ and is given by,

$$\mathrm{SD} = \sqrt{\sum_j \left( X_j^{exp} - X_j^{pred} \right)^2 / N} \tag{10}$$

- *Average absolute error* (AAE): This is the measure of deviation of predicted property values from the experimentally measured property values and is given by,

$$\mathrm{AAE} = \frac{1}{N} \sum_j \left| X_j^{exp} - X_j^{pred} \right| \tag{11}$$

- *Average relative error* (ARE): This provides an average of relative error calculated with respect to the experimentally measured property values and is given by,

$$\mathrm{ARE} = \frac{1}{N} \sum_j \left| \left( X_j^{exp} - X_j^{pred} \right) / X_j^{exp} \right| \times 100 \tag{12}$$

- *Coefficient of determination* ($R^2$): This parameter provides information about the goodness of model fit. An $R^2$ close to 1.0 indicates that the experimental data used in the regression have been fitted to a good accuracy. It is calculated using,

$$\mathrm{R}^2 = 1 - \left[ \sum_j \left( X_j^{exp} - X_j^{pred} \right)^2 / \sum_j \left( X_j^{exp} - \mu \right)^2 \right] \tag{13}$$

The indicators SD, AAE, ARE, and $R^2$ provide measures of quality (reliability) of property prediction models on a global basis. However, it is important that the information of uncertainties of estimated values also be made available to the user in order to provide confidence in the estimated property values and hence in the design of sustainable processes.

# RESULTS

In this section, the selection of suitable property models for modeling environment-related properties and the performance statistics for the developed property models are discussed. The results are presented for the following models:

- MG method based property models analyzed using step-wise regression method
- MG method based property models analyzed using simultaneous regression method
- CI method based property models

**Selection of Suitable Property Models for Environment-related Properties.** In this work, the basis for selecting an appropriate property model for the environment-related property has been the study of behaviour of that property of certain class of chemicals with increasing carbon number. This is illustrated for the case of $LC_{50}$(FM). Figure 1 shows a plot of primary alcohols with increasing carbon number (propyl alcohol, butyl alcohol, and so on) versus their experimental values of –Log $LC_{50}$(FM). It can be seen that this plot is almost a straight line (with $R^2$ value of 0.99) suggesting that the property $LC_{50}$ (FM) can be modelled using a linear model of the form $-\text{Log}LC_{50}(\text{FM}) + \text{Constant} = \sum_i N_i C_i + w\sum_j M_j D_j + z\sum_k E_k O_k$ . Similar analysis have been performed (not shown in this paper) to obtain a suitable form of the property model for other environment-related properties with the objective of providing an accurate and reliable property estimation of environment-related properties.

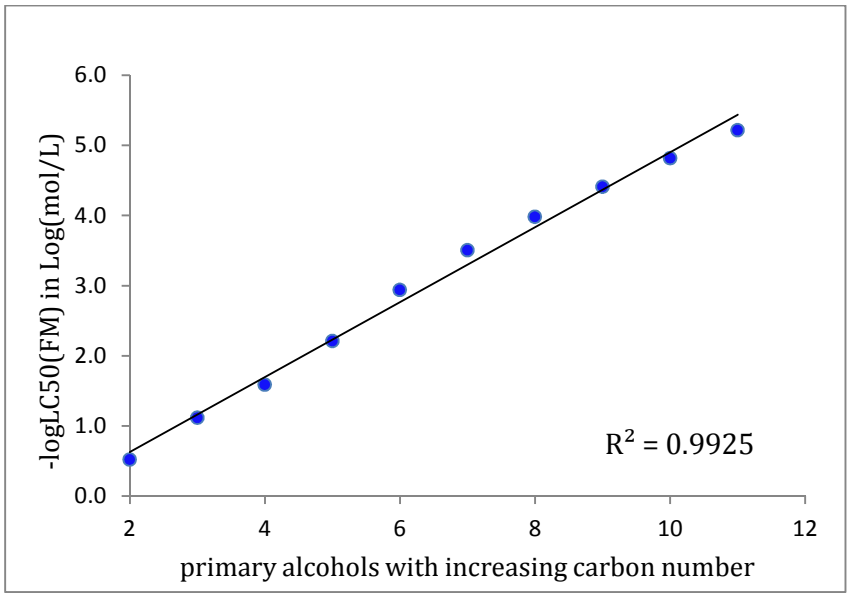



**Figure 1.** Plot of primary alcohols versus their experimental values of $-LogLC_{50}(FM)$

**Model Performance.** The model performance statistics for property models analysed using the step-wise regression method are provided in Table 2. The model performance statistics for properties analysed using the simultaneous regression method are given in Table S1 in the supporting information which can be downloaded from the following link: http://www.capec.kt.dtu.dk/documents/environment_related_properties/supporting_information.pdf. In Table 2, N is the number of experimental data-points considered in the regression and *v* is the degrees of freedom and is obtained by subtracting number of estimated model parameters from N. $P_{rc(\pm 1\%)}$, $P_{rc(\pm 5\%)}$, and $P_{rc(\pm 10\%)}$ represents the percentage of the experimental data-points (N) found within ± 1%, ± 5%, and ± 10% relative error range respectively. For property models analysed using step-wise regression method, the results for $R^2$, SD, AAE and ARE have been obtained after third-level estimation; hence, they represent the global results of the three sequential approximations. The residuals ($X^{exp}$-$X^{pred}$) for data-points considered in the regression are plotted in the form of residual distribution plots and are included in Table 2 and Table S1 (note that Table S1 is included as supporting information). For most of the property models (except for ozone depletion potential and

acidification potential) the residuals followed a normal distribution curve suggesting that the assumption of normal distribution of random errors is valid behind the followed approach. The model performance statistics for property models analysed using the CI method are provided in Table 3. These CI models have been employed together with the GC method for creating the missing groups and predicting their contributions through the regressed contributions of connectivity indices as suggested by Gani et al.[14] This feature makes it possible to predict environment-related properties of organic chemicals for which neither experimental data nor the GC-property model parameters are available. The property models developed based on the CI method (see Table 3) have reasonable model performance statistics. High accuracy in the prediction of environment-related properties cannot be expected from this model since, only a few parameters are involved to represent large data-set of chemicals. Greater accuracy can be obtained by adding higher-order connectivity indices. However, the main objective of analyzing CI models in this work is to obtain the missing group contributions, for which only the first two connectivity indices should be sufficient.[14] Hukkerikar et al.[24] discussed the effect of quantity of experimental data on the quality of parameter estimation and illustrated that by including all of the available experimental data of the property in the regression it is possible to improve the predictive capability and application range of the property model. Therefore, in this work we have considered all of the available experimental data of properties of chemicals for modeling environment-related properties. To illustrate this point, we have considered here an analysis of property model for oral rat $LD_{50}$, fathead minnow 96-hr $LC_{50}$, and emission to continental rural air (carcinogenic ($ERA_C$) and non-carcinogenic ($ERA_{NC}$)). The whole experimental data-sets of these properties (5995 data-points for oral rat $LD_{50}$, 809 data-points for fathead minnow 96-hr $LC_{50}$, 456 data-points for emission to continental rural air (carcinogenic ($ERA_C$), and 341 data-points for emission to continental rural air (non-carcinogenic

($ERA_{NC}$))) is divided randomly in 5 subsets (A, B, C, D, and E) of equal size. The property model is trained on 4 subsets (using simultaneous regression method) and 1 subset is used for testing purpose. This procedure is repeated 5 times so that all subsets are used for testing purpose. The results in terms of SD, AAE, and ARE for training sets and for test sets is presented in Table 4(a) for oral rat $LD_{50}$, Table 4(b) for fathead minnow 96-hr $LC_{50}$, Table 4(c) for emission to continental rural air (carcinogenic ($ERA_C$) and in Table 4(d) for emission to continental rural air (non-carcinogenic ($ERA_{NC}$)). The *MSECV*, which is mean squared error of cross-validation (Mevik and Cederkvist[28]) calculated using eq. (14) is also given in Tables 4(a)-4(d).

$$MSECV = \frac{1}{N_L}\sum_{k=1}^{K}\sum_{j\in L_k}\left(X_j^{exp} - X_j^{pred}\right)^2 \qquad (14)$$

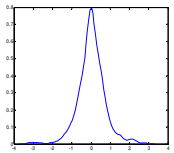
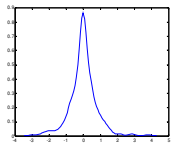
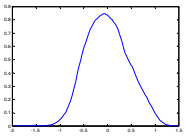
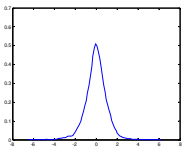
Where, $N_L$ is the number of data-points in the training set, $K$ = number of subsets (5 in this analysis), and $L_k$ is the number of data-points in the subsets.

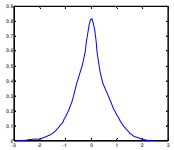
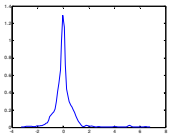
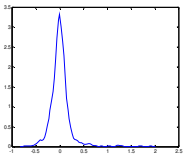
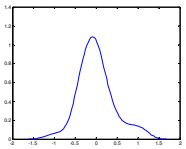
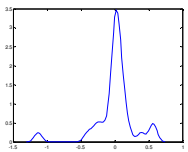
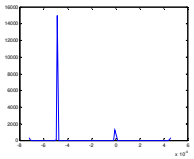
From Table 4(a), comparison of the model performance for training sets and test sets show that the predictive capability of the model for oral rat $LD_{50}$ is fairly good. This is mainly due to the large amount of available experimental data of oral rat $LD_{50}$ for the training purpose. For test sets, if we compare the SD, AAE, and ARE values calculated using the parameters obtained by regressing training set with those that are calculated using the parameters obtained by regression of the whole data-set, we find that better model performance statistics (lower SD, lower AAE, and lower ARE) is obtained when we use model parameters that are estimated using all of the experimental data-points in the regression.

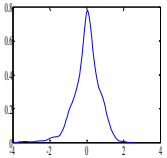
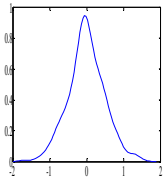
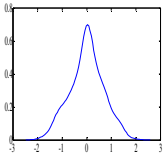
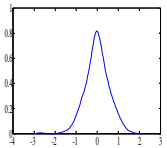
For fathead minnow 96-hr $LC_{50}$, emission to continental rural air (carcinogenic ($ERA_C$)), and emission to continental rural air (non-carcinogenic ($ERA_{NC}$)) it can be seen from Tables 4(b)-4(d) that the model performance for test sets is poor as compared to those of training sets

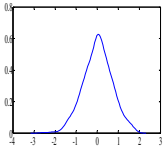
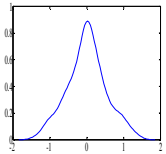
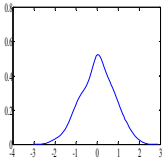
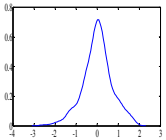
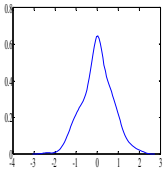
and this is due to the small amount of available experimental data of these properties for the training purpose. For these properties, it can be observed that the SD, AAE, and ARE values for test sets calculated using the model parameters as obtained by regression of the whole data-set are much better than those that are calculated using the parameters estimated using the training set indicating the importance of considering all of the available experimental data-points for the regression purpose. To sum up, this analysis shows both the robustness of the approach and the predictive capability of the developed models for estimating environmental related properties.

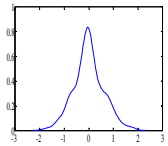
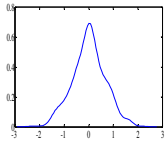
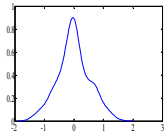
**Table 2. Performance of MG Method Based Property Models Analysed Using Step-wise Regression Method**

| sl. no. | property | L.H.S. of MG method based property prediction model | MG group-contribution model $f(X)=\sum_i N_i C_i + \sum_j M_j D_j + \sum_k E_k O_k$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $f(X)$ | N | $v$ | $R^2$ | residual distribution plot | $P_{rc}$ ($\pm$ 1%) | $P_{rc}$ ($\pm$ 5%) | $P_{rc}$ ($\pm$ 10%) | SD | AAE | ARE [a] |
| 1 | fathead minnow 96-hr $LC_{50}$ ($LC_{50}$ (FM)) in mol/lit | $-\mathrm{LogLC_{50}(FM)} + \mathrm{FM_0}$ | 809 | 541 | 0.78 |  | 8.53 | 31.52 | 54.02 | 0.69 | 0.48 | 21.56 |
| 2 | *daphnia magna* 48-hr $LC_{50}$ ($LC_{50}$ (DM)) in mol/lit | $-\mathrm{LogLC_{50}(DM)} + \mathrm{DM_0}$ | 320 | 124 | 0.82 |  | 16.25 | 39.06 | 62.50 | 0.74 | 0.49 | 16.16 |
| 3 | oral rat $LD_{50}$ ($LD_{50}$) in mol/kg | $-\mathrm{LogLD50} - \mathrm{A_{LD50}} - \mathrm{B_{LD50}MW}$ | 5995 | 5617 | 0.73 |  | 1.52 | 6.92 | 13.61 | 0.43 | 0.35 | 16.40 |
| 4 | aqueous solubility ($LogW_S$) in gm/lit | $\mathrm{LogW}_S - \mathrm{A_{W_s}} - \mathrm{B_{W_s}MW}$ | 4681 | 4311 | 0.78 |  | 3.12 | 14.36 | 28.63 | 0.99 | 0.73 | ---- |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | bioconcentra -tion factor (BCF) | LogBCF | 662 | 423 | 0.78 |  | 8.91 | 19.49 | 30.82 | 0.63 | 0.47 | ---- |
| 6 | permissible exposure limit (OSHA-TWA) in mol/lit | -LogPEL | 425 | 239 | 0.74 |  | 16.71 | 39.53 | 60.24 | 0.78 | 0.44 | 12.61 |
| 7 | photochemic al oxidation potential (PCO) | -LogPCO | 639 | 488 | 0.83 |  | 6.42 | 16.9 | 26.30 | 0.22 | 0.13 | 8.37 |
| 8 | global warming potential (GWP) | LogGWP | 51 | 31 | 0.87 |  | 15.69 | 37.25 | 56.86 | 0.41 | 0.29 | 11.57 |
| 9 | ozone depletion potential (ODP) | LogODP | 28 | 12 | 0.89 |  | 17.86 | 21.4 | 28.5 | 0.30 | 0.16 | ---- |
| 10 | acidification potential (ODP) | LogAP | 10 | 1 | 1.0 |  | 100.0 | -- | -- | 3.4E-04 | 2.1E-4 | ---- |

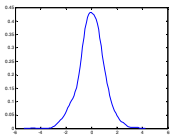
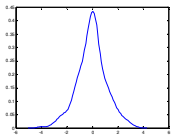
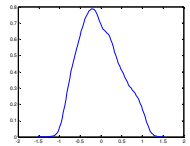
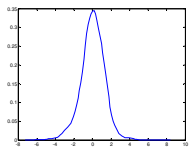
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | emission to urban air ($EUA_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(EUA_C)+A_{EUA_C}$ | 456 | 214 | 0.70 |  | 16.23 | 40.13 | 63.60 | 0.70 | 0.50 | 10.61 |
| 12 | emission to urban air ($EUA_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(EUA_{NC})+A_{EUA_{NC}}$ | 341 | 128 | 0.79 |  | 12.90 | 47.80 | 76.50 | 0.49 | 0.37 | 6.80 |
| 13 | emission to continental rural air ($ERA_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ERA_C)+A_{ERA_C}$ | 470 | 229 | 0.75 |  | 15.74 | 39.15 | 64.89 | 0.67 | 0.51 | 8.88 |
| 14 | emission to continental rural air ($ERA_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(ERA_{NC})+A_{ERA_{NC}}$ | 349 | 134 | 0.80 |  | 13.18 | 46.13 | 75.07 | 0.55 | 0.42 | 7.25 |

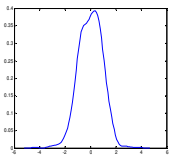
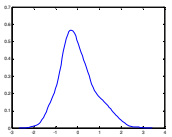
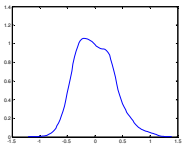
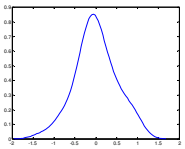
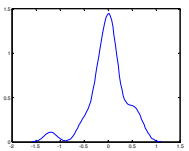
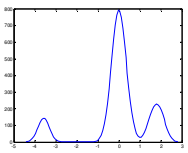
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | emission to continental fresh water ($EFW_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(EFW_C)+A_{EFW_C}$ | 472 | 230 | 0.75 |  | 13.77 | 31.77 | 60.16 | 0.67 | 0.52 | 11.26 |
| 16 | emission to continental fresh water ($EFW_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(EFW_{NC})+A_{EFW_{NC}}$ | 345 | 131 | 0.83 |  | 13.33 | 44.63 | 67.82 | 0.52 | 0.40 | 8.15 |
| 17 | emission to continental sea water ($ESW_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ESW_C)+A_{ESW_C}$ | 477 | 235 | 0.81 |  | 15.30 | 37.94 | 67.71 | 0.79 | 0.61 | 8.69 |
| 18 | emission to continental sea water ($ESW_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(ESW_{NC})+A_{ESW_{NC}}$ | 360 | 146 | 0.85 |  | 14.16 | 46.38 | 72.22 | 0.69 | 0.51 | 8.41 |
| 19 | emission to continental natural soil ($ENS_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ENS_C)+A_{ENS_C}$ | 472 | 231 | 0.76 |  | 13.98 | 39.61 | 63.55 | 0.72 | 0.55 | 9.28 |

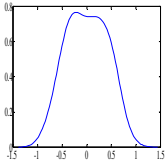
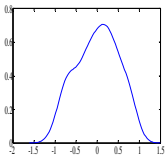
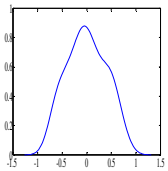
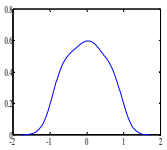
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | emission to continental natural soil ($ENS_{NC}$) in cases/kg emitted (non-carcinogenic) | $-Log(ENS_{NC})+A_{ENS_{NC}}$ | 362 | 148 | 0.79 |  | 14.91 | 48.06 | 71.27 | 0.61 | 0.46 | 7.27 |
| 21 | emission to continental agri-cultural soil ($EAS_C$) in cases/kg emitted (carcinogenic) | $-Log(EAS_C)+A_{EAS_C}$ | 470 | 228 | 0.75 |  | 13.61 | 41.06 | 65.74 | 0.67 | 0.51 | 9.36 |
| 22 | emission to continental agri-cultural soil ($EAS_{NC}$) in cases/kg emitted (non-carcinogenic) | $-Log(EAS_{NC})+A_{EAS_{NC}}$ | 352 | 138 | 0.80 |  | 16.19 | 48.29 | 74.71 | 0.54 | 0.41 | 6.92 |

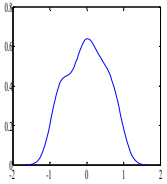
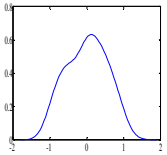
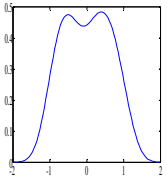
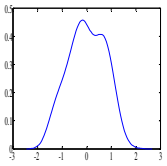
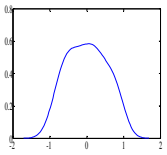
[a] ARE is not defined for LogWs, BCF, ODP and AP since these properties have both positive and negative values.

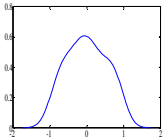
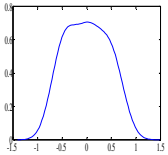
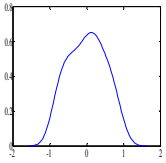
**Table 3. Performance of CI Method Based Property Models**

| sl. no. | property | L.H.S. of CI method based property prediction model | atom connectivity index (CI) model $f(X)=\sum_i a_i A_i + b\left({}^{v}\chi^{0}\right)+2c\left({}^{v}\chi^{1}\right)+d$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $f(X)$ | N | $v$ | $R^2$ | residual distribution plot | $P_{rc}$ ($\pm$ 1%) | $P_{rc}$ ($\pm$ 5%) | $P_{rc}$ ($\pm$ 10%) | SD | AAE | ARE [a] |
| 1 | fathead minnow 96-hr $LC_{50}$ ($LC_{50}$ (FM)) in mol/lit | $-\mathrm{LogLC}_{50}(\mathrm{FM})+\mathrm{FM}_0$ | 809 | 796 | 0.56 |  | 3.96 | 16.70 | 34.20 | 0.98 | 0.75 | 40.47 |
| 2 | *daphnia magna* 48-hr $LC_{50}$ ($LC_{50}$ (DM)) in mol/lit | $-\mathrm{LogLC}_{50}(\mathrm{DM})+\mathrm{DM}_0$ | 320 | 307 | 0.58 |  | 5.0 | 22.81 | 40.94 | 1.14 | 0.85 | 35.21 |
| 3 | oral rat $LD_{50}$ ($LD_{50}$) in mol/kg | $-\mathrm{LogLD50}-A_{\mathrm{LD50}}-B_{\mathrm{LD50}}\mathrm{MW}$ | 5662 | 5647 | 0.60 |  | 1.02 | 5.35 | 11.48 | 0.48 | 0.40 | 18.49 |
| 4 | aqueous solubility ($LogW_S$) in gm/lit | $\log(W_S)-A_{W_S}-B_{W_s}\mathrm{MW}$ | 4681 | 4676 | 0.62 |  | 2.22 | 9.98 | 19.80 | 1.29 | 0.98 | ---- |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | bioconcentra -tion factor (BCF) | LogBCF | 662 | 648 | 0.53 |  | 1.66 | 6.19 | 12.54 | 0.92 | 0.74 | ---- |
| 6 | permissible exposure limit (PEL) in mol/lit | -LogPEL | 411 | 397 | 0.64 |  | 4.87 | 16.79 | 33.09 | 0.78 | 0.61 | 20.10 |
| 7 | photochemic al oxidation potential (PCO) | -LogPCO | 621 | 607 | 0.51 |  | 1.61 | 4.83 | 8.05 | 0.33 | 0.27 | 16.65 |
| 8 | global warming potential (GWP) | LogGWP | 51 | 37 | 0.83 |  | 9.80 | 31.37 | 50.98 | 0.48 | 0.36 | 15.52 |
| 9 | ozone depletion potential (ODP) | LogODP | 28 | 14 | 0.83 |  | 7.14 | 10.71 | 14.30 | 0.37 | 0.25 | ---- |
| 10 | acidification potential (ODP) | LogAP | 10 | 1 | 1.00 |  | 70.0 | 100.0 | -- | 0.0014 | 802E-04 | -- |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | emission to urban air ($EUA_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(EUA_C)+A_{EUA_C}$ | 232 | 220 | 0.66 |  | 5.17 | 40.08 | 83.18 | 0.40 | 0.34 | 6.36 |
| 12 | emission to urban air ($EUA_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(EUA_{NC})+A_{EUA_{NC}}$ | 259 | 247 | 0.66 |  | 7.72 | 40.92 | 69.11 | 0.49 | 0.41 | 7.50 |
| 13 | emission to continental rural air ($ERA_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ERA_C)+A_{ERA_C}$ | 226 | 214 | 0.79 |  | 11.94 | 49.55 | 88.05 | 0.39 | 0.32 | 5.43 |
| 14 | emission to continental rural air ($ERA_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(ERA_{NC})+A_{ERA_{NC}}$ | 257 | 245 | 0.74 |  | 7.78 | 39.69 | 68.48 | 0.53 | 0.44 | 7.57 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | emission to continental fresh water ($EFW_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(EFW_C)+A_{EFW_C}$ | 286 | 274 | 0.65 |  | 7.34 | 36.36 | 61.53 | 0.52 | 0.44 | 8.51 |
| 16 | emission to continental fresh water ($EFW_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(EFW_{NC})+A_{EFW_{NC}}$ | 259 | 247 | 0.70 |  | 9.26 | 33.59 | 60.61 | 0.54 | 0.44 | 9.02 |
| 17 | emission to continental sea water ($ESW_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ESW_C)+A_{ESW_C}$ | 286 | 274 | 0.78 |  | 4.89 | 35.31 | 75.17 | 0.62 | 0.54 | 7.20 |
| 18 | emission to continental sea water ($ESW_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(ESW_{NC})+A_{ESW_{NC}}$ | 291 | 279 | 0.77 |  | 5.84 | 32.30 | 65.63 | 0.72 | 0.60 | 8.76 |
| 19 | emission to continental natural soil ($ENS_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ENS_C)+A_{ENS_C}$ | 285 | 273 | 0.61 |  | 6.66 | 38.59 | 74.38 | 0.52 | 0.44 | 6.89 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | emission to continental natural soil ($ENS_{NC}$) in cases/kg emitted (non-carcinogenic) | $-Log(ENS_{NC})+A_{ENS_{NC}}$ | 247 | 235 | 0.70 |  | 9.31 | 39.67 | 72.06 | 0.53 | 0.45 | 7.08 |
| 21 | emission to continental agri-cultural soil ($EAS_C$) in cases/kg emitted (carcinogenic) | $-Log(EAS_C)+A_{EAS_C}$ | 240 | 228 | 0.68 |  | 8.33 | 42.50 | 88.33 | 0.42 | 0.36 | 5.76 |
| 22 | emission to continental agri-cultural soil ($EAS_{NC}$) in cases/kg emitted (non-carcinogenic) | $-Log(EAS_{NC})+A_{EAS_{NC}}$ | 247 | 235 | 0.70 |  | 8.50 | 40.89 | 74.08 | 0.49 | 0.42 | 6.94 |

[a] ARE is not defined for LogWs, BCF, ODP and AP since these properties have both positive and negative values.

**Table 4(a). Performance of Model for Oral Rat $LD_{50}$ Based on Different Combinations of Training Sets and Test Sets**

| datasets used for | | model performance statistics for training set | | | | model performance statistics for test set using the parameters estimated from regression of the training set | | | model performance statistics for test set using the parameters estimated from regression of the whole data-set (containing 5995 data-points) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| training purpose | testing purpose | *MSECV* | SD Log(mol/kg) | AAE Log(mol/kg) | ARE % | SD Log(mol/kg) | AAE Log(mol/kg) | ARE % | SD Log(mol/kg) | AAE Log(mol/kg) | ARE % |
| A, B, C, D | E | 0.1812 | 0.4257 | 0.3479 | 15.97 | 0.4628 | 0.3732 | 17.32 | 0.4220 | 0.3424 | 15.91 |
| A, B, C, E | D | 0.1796 | 0.4238 | 0.3456 | 15.90 | 0.4755 | 0.3839 | 17.72 | 0.4287 | 0.3506 | 16.21 |
| A, B, D, E | C | 0.1805 | 0.4248 | 0.3462 | 15.97 | 0.4754 | 0.3823 | 17.23 | 0.4251 | 0.3500 | 15.90 |
| A, C, D, E | B | 0.1788 | 0.4229 | 0.3449 | 15.89 | 0.4677 | 0.3813 | 17.20 | 0.4338 | 0.3536 | 16.00 |
| B, C, D, E | A | 0.1794 | 0.4236 | 0.3455 | 15.86 | 0.4694 | 0.3848 | 17.93 | 0.4302 | 0.3532 | 16.46 |
| average performance | | 0.1799 | 0.4241 | 0.3460 | 15.91 | 0.4701 | 0.3811 | 17.48 | 0.4279 | 0.3499 | 16.09 |

**Table 4(b). Performance of Model for Fathead Minnow 96-hr $LC_{50}$ Based on Different Combinations of Training Sets and Test Sets**

| datasets used for | | model performance statistics for training set | | | | model performance statistics for test set using the parameters estimated from regression of the training set | | | model performance statistics for test set using the parameters estimated from regression of the whole data-set (containing 809 data-points) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| training purpose | testing purpose | *MSECV* | SD Log(mol/kg) | AAE Log(mol/kg) | ARE % | SD Log(mol/kg) | AAE Log(mol/kg) | ARE % | SD Log(mol/kg) | AAE Log(mol/kg) | ARE % |
| A, B, C, D | E | 0.3400 | 0.5831 | 0.4015 | 19.68 | 1.3753 | 0.8615 | 27.04 | 0.6732 | 0.4786 | 15.68 |
| A, B, C, E | D | 0.3339 | 0.5778 | 0.3991 | 19.10 | 1.3944 | 0.9325 | 28.91 | 0.6854 | 0.4778 | 17.65 |
| A, B, D, E | C | 0.3624 | 0.6020 | 0.4237 | 14.99 | 1.3517 | 0.9127 | 47.19 | 0.6581 | 0.4802 | 33.19 |
| A, C, D, E | B | 0.3645 | 0.6037 | 0.4201 | 20.39 | 1.4857 | 0.9072 | 28.26 | 0.6399 | 0.4654 | 14.62 |
| B, C, D, E | A | 0.3453 | 0.5876 | 0.4142 | 17.62 | 1.5178 | 0.9710 | 35.12 | 0.6722 | 0.4831 | 21.19 |
| average performance | | 0.3492 | 0.5908 | 0.4117 | 18.35 | 1.4249 | 0.9169 | 33.30 | 0.6657 | 0.4770 | 20.47 |

**Table 4(c). Performance of Model for Emission to Urban Air (Carcinogenic) Based on Different Combinations of Training Sets and Test Sets**

| datasets used for | | model performance statistics for training set | | | | model performance statistics for test set using the parameters estimated from regression of the training set | | | model performance statistics for test set using parameters estimated from regression of the whole data-set (containing 456 data-points) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| training purpose | testing purpose | *MSECV* | SD Cases/kg emitted | AAE Cases/kg emitted | ARE % | SD Cases/kg emitted | AAE Cases/kg emitted | ARE % | SD Cases/kg emitted | AAE Cases/kg emitted | ARE % |
| A, B, C, D | E | 0.2050 | 0.4528 | 0.3024 | 5.86 | 1.8293 | 1.1974 | 27.82 | 0.5386 | 0.4135 | 9.16 |
| A, B, C, E | D | 0.2206 | 0.4697 | 0.3252 | 6.55 | 1.4664 | 0.9823 | 17.63 | 0.4766 | 0.3534 | 6.60 |
| A, B, D, E | C | 0.1675 | 0.4093 | 0.2871 | 5.52 | 1.6849 | 1.2268 | 25.79 | 0.6160 | 0.4165 | 9.52 |
| A, C, D, E | B | 0.2111 | 0.4595 | 0.3052 | 6.42 | 2.0921 | 1.3187 | 24.12 | 0.4854 | 0.3480 | 6.33 |
| B, C, D, E | A | 0.2182 | 0.4671 | 0.3267 | 6.63 | 1.3597 | 1.0115 | 19.55 | 0.4572 | 0.3478 | 6.55 |
| average performance | | 0.2045 | 0.4517 | 0.3093 | 6.19 | 1.6865 | 1.1473 | 22.98 | 0.4713 | 0.3479 | 6.44 |

**Table 4(d). Performance of Model for Emission to Urban Air (Non-Carcinogenic) Based on Different Combinations of Training Sets and Test Sets**

| datasets used for | | model performance statistics for training set | | | | model performance statistics for test set using the parameters estimated from regression of the training set | | | model performance statistics for test set using parameters estimated from regression of the whole data-set (containing 341 data-points) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| training purpose | testing purpose | *MSECV* | SD Cases/kg emitted | AAE Cases/kg emitted | ARE % | SD Cases/kg emitted | AAE Cases/kg emitted | ARE % | SD Cases/kg emitted | AAE Cases/kg emitted | ARE % |
| A, B, C, D | E | 0.0655 | 0.2560 | 0.1716 | 3.14 | 2.3797 | 1.4854 | 26.38 | 0.3872 | 0.2801 | 5.03 |
| A, B, C, E | D | 0.0997 | 0.3157 | 0.2217 | 4.12 | 1.7248 | 1.0980 | 19.22 | 0.3693 | 0.2708 | 4.89 |
| A, B, D, E | C | 0.0831 | 0.2882 | 0.1894 | 3.50 | 2.4662 | 1.5963 | 30.72 | 0.3684 | 0.2615 | 5.01 |
| A, C, D, E | B | 0.0846 | 0.2909 | 0.1936 | 3.51 | 3.0908 | 1.6401 | 29.28 | 0.3980 | 0.3007 | 5.62 |
| B, C, D, E | A | 0.1097 | 0.3313 | 0.2289 | 4.23 | 8.7061 | 3.9458 | 80.85 | 0.3096 | 0.2166 | 3.80 |
| average performance | | 0.0885 | 0.2964 | 0.2010 | 3.69 | 3.6735 | 1.9531 | 37.28 | 0.3665 | 0.2659 | 4.87 |

Marrero and Gani[13] reported SD, AAE, and $R^2$ values for the GC model for $LogW_s$ as 0.55, 0.46 and 0.93 respectively. In their analysis, the number of estimated model parameters (groups) are 155 first order groups, 99 second order groups, and 48 third-order groups (that is, total 302 groups estimated out of 424 groups). Referring to the Table 2, it can be seen that the property model for $LogW_s$ has SD, AAE, and $R^2$ values of 0.99, 0.73, and 0.78 respectively. In this work, the number of estimated groups are 197 first order groups, 124 second order groups, and 57 third-order groups (total 378 groups estimated out of 424 groups). It is to be noted that in the present work, a much larger data-set (4681 data-points as compared to 2087 data-points used by Marrero and Gani[13]) of $LogW_s$ comprising complex and polyfunctional environment-related chemicals is used in the regression, which makes it possible to estimate larger number of model parameters thereby contributing to improved application range of the property model for $LogW_s$. A similar note can be made for the case of property model for $LC_{50}$(FM). The developed property model for $LC_{50}$(FM) has SD, AAE, and $R^2$ values of 0.69, 0.48, and 0.78 respectively. Martin and Young[15] reported SD and $R^2$ values for their GC model for $LC_{50}$(FM) as 0.37 and 0.91 respectively. The use of the large data-set for $LC_{50}$ (FM) allows estimation of a large number of model parameters which in turn allows one to estimate $LC_{50}$ (FM) for a wide range of organic chemicals. For the property $LC_{50}$(DM), the model performance statistics are similar to that of $LC_{50}$(FM) model. The developed property model for $LD_{50}$ (using a data-set of 5995 chemicals) has reasonably good performance statistics with SD, AAE, and $R^2$ values as 0.43, 0.35 and 0.73 respectively. Several estimation methods based on the QSAR approach have been reported in the literature that uses other properties such as $LC_{50}$(DM) as an input to their estimation method to estimate $LD_{50}$. Also, these methods have been developed based on relatively smaller data-sets (with few hundreds of chemicals in the data-set) of chemicals. The application of such methods is restricted by the availability of the experimental data of the needed input

properties for their estimation. A similar issue is associated with the estimation methods for BCF requiring additional inputs such as the octanol/water partition coefficient. In this work, the developed property model for BCF has SD, AAE, and $R^2$ values of 0.63, 0.47, and 0.78 respectively. Note, that the developed property models for $LD_{50}$ and for BCF only require the molecular structure of the chemical for the property estimation. For properties GWP, ODP, and AP the number of experimental data points used in the regression are smaller as compared to other properties analysed in this work. However, it can be noted that these properties belong to a particular class of chemicals (for example, global warming potential and ozone depletion potential properties involve halogenated chemicals, acidification potential property involves nitrogenated chemicals); hence, even though the experimental data-sets are smaller, the models for these properties are able to provide estimation of these properties with good accuracy. The model performance statistics for the remaining properties namely, $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$ $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_{C,}$ and $EAS_{NC}$ show that the experimental data have been fitted to a good degree of accuracy. The estimation of these properties is based exclusively on the molecular structure of the chemical and allows the user to calculate Human Toxicity Potential (HTP) [8] (which is needed to perform life cycle impact assessment of the product) thus increasing the application range of the USEtox™ model to a wide range of chemicals.

The variables $FM_0$,$DM_0$, $A_{LogWs}$, $B_{LogWs}$ $A_{LD50}$, $B_{LD50}$, $A_{EUAc}$, $A_{EUAnc}$, $A_{ERAc}$, $A_{ERAnc}$, $A_{EFWc}$, $A_{EFWnc}$, $A_{ESWc}$, $A_{ESWnc}$, $A_{ENSc}$, $A_{ENSnc}$, $A_{EASc}$, $A_{EASnc}$ as defined in the functional forms, $f(X)$ given in Tables 2-3 are additional adjustable parameters of property prediction models. The values of these parameters are listed in Table 5. The total list of groups and their contributions $C_i$, $D_j$, and $E_k$ for the 22 environment-related properties analysed in this work are given in the supporting information (see Tables S2-S4 for MG method based models analysed using step-wise regression method, and Tables S5-S7 for MG method based models analysed using simultaneous regression method). The list of atoms, their contributions $a_i$, adjustable model parameters ($b$ and $c$), and the universal parameter $d$ for CI method based property prediction models are given in the supporting information (see Table S8). The covariance matrix computed using eq. (6) for each property prediction model analyzed using the MG method (for models with step-wise regression method and simultaneous regression method) and using the CI method is available upon request from the authors. The developed models for environment-related properties (for $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), PCO, GWP, ODP, and AP) have been implemented in ProPred, a property estimation toolbox of ICAS ® (Integrated Computer Aided System[29]) software developed by CAPEC, DTU.

**Table 5. Values of Universal Constants (Additional Adjustable Parameters)** [a]

| universal constants | value (step-wise method) | value (simultanous method) |
|---|---|---|
| $FM_0$ | 2.1949 | 2.1841 |
| $DM_0$ | 2.9717 | 3.5907 |
| $A_{LogWs}$ | 4.5484 | 4.3098 |
| $B_{LogWs}$ | 0.3411 | 0.3404 |
| $A_{LD50}$ | 1.9372 | 1.9372 |
| $B_{LD50}$ | 0.0016 | 0.0016 |
| $A_{EUAC}$ | 5.2801 | 5.22536 |
| $A_{EUANC}$ | 6.8181 | 7.06605 |
| $A_{ERAC}$ | 6.5561 | 6.68611 |
| $A_{ERANC}$ | 7.5541 | 9.53269 |
| $A_{EFWC}$ | 5.6726 | 5.0706 |
| $A_{EFWNC}$ | 6.4429 | 7.33378 |
| $A_{ESWC}$ | 8.3962 | 9.33319 |
| $A_{ESWNC}$ | 8.6360 | 10.0724 |
| $A_{ENSC}$ | 6.4837 | 5.93334 |
| $A_{ENSNC}$ | 7.0265 | 6.4159 |
| $A_{EASC}$ | 6.2913 | 5.48504 |
| $A_{EASNC}$ | 6.9723 | 6.06003 |

[a] values of universal constants for the CI models are the same as those based on the step-wise method.

**Application of the Developed Property Models for the Estimation of Environment-Related Properties.** The application of the developed property models to estimate properties of environment-related properties and to quantify the uncertainties of the estimated property values is illustrated by considering predictions of $LogW_s$ (using model parameters obtained from simultaneous regression method) for the chemical, Benzo[a]pyrene, (CAS No. 50-32-8) which is a polycyclic aromatic hydrocarbon and is highly carcinogenic. The experimentally measured value of $LogW_s$ (mg/L) for Benzo[a]pyrene is -2.79. Table 6 provides information of first-order, second-order, and third-order groups used to represent Benzo[a]pyrene, their frequency (that is, occurrences in the structure) and the contributions for each group ($LogW_{s1i}$, $LogW_{s2j}$, and $LogW_{s3k}$) taken from Tables S5-S7 given in the supporting

information. Using this information and the universal constants of the property model for $LogW_s$, the value of $LogW_s$ for Benzo[a]pyrene was estimated as -2.64 (with absolute error = |-2.79 – (-2.64)| = 0.15).

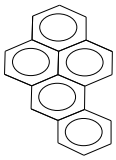
**Table 6. Estimation of $LogW_s$ of Benzo[a]pyrene**

Benzo[a]pyrene

CAS No. 50-32-8

molecular formula: $C_{20}H_{12}$

molecular structure



| **first-order groups** | **occurrences** | **contribution** |
|---|---|---|
| aCH | 12 | -4.5565 |
| aC fused with aromatic ring | 8 | -4.7557 |
| **second-order groups** | **occurrences** | **contribution** |
| No second-order groups are involved | | |
| **third-order groups** | **occurrences** | **contribution** |
| AROM.FUSED[2] | 1 | -0.0759 |
| AROM.FUSED[3] | 1 | -0.1255 |
| AROM.FUSED[4p] | 2 | 0.0500 |

$$\mathrm{LogW_s} = \mathrm{A_{LogWs}} + (\mathrm{B_{LogWs}} \cdot \mathrm{MW}) + \sum_i N_i C_i + w \sum_j M_j D_j + z \sum_k E_k O_k = -2.64$$

note: estimated value of $LogW_s$ using Marrero and Gani (2002) method = -2.20

As a next step, the uncertainty of the estimated $LogW_s$ is quantified. For this purpose, information of covariance $COV(\boldsymbol{P}^*)$ of the involved groups and the universal constants $A_{LogWs}$ and $B_{LogWs}$ and also the local sensitivity $J(\boldsymbol{P}^*)$ of the $LogW_s$ model is needed. The covariance of the involved groups (as listed in Table 7) and universal constants $A_{LogWs}$ and $B_{LogWs}$ was noted from the overall covariance matrix for all the groups of the $LogW_s$ model analysed using simultaneous regression method. In Table 7, only lower triangular elements are shown since the upper triangular matrix elements are identical to the lower ones. Table 8 lists the local sensitivity of the $LogW_s$ model with respect to the model parameters (for contributions listed in Table 6 and universal constants $A_{LogWs}$ and $B_{LogWs}$).

**Table 7. Covariance Matrix $COV(\boldsymbol{P}^*)$ with Dimensions (7×7)**

| | $A_{LogWs}$ | $B_{LogWs}$ | aCH | aC | AROM.FUSED[2] | AROM.FUSED[3] | AROM.FUSED[4p] |
|---|---|---|---|---|---|---|---|
| $A_{LogWs}$ | 0.0154 | | | | | | |
| $B_{LogWs}$ | -1.28E-07 | 4.97E-07 | | | | | |
| aCH | -0.0025 | -8.1E-06 | 6.71E-04 | | | | |
| aC | -0.002 | -4.7E-06 | -3.7E-04 | 0.0048 | | | |
| AROM.FUSED[2] | 7.7E-05 | -4.3E-06 | -5.9E-04 | -0.0047 | 0.0113 | | |
| AROM.FUSED[3] | -0.0013 | -2.4E-06 | -8.9E-06 | -0.0084 | 0.0111 | 0.0375 | |
| AROM.FUSED[4p] | -4.5E-04 | 1.6E-06 | 8.25E-06 | -0.0092 | 0.009 | 0.0136 | 0.0283 |

**Table 8. Local Sensitivity $J(\boldsymbol{P}^*)$ with Dimensions (1×7) of $LogW_s$ Model with Respect to the Model Parameters**

| $\delta LogWs/\delta A_{LogWs}$ | $\delta LogWs/\delta B_{LogWs}$ | $\delta LogWs/\delta$aCH | $\delta LogWs/\delta$aC | $\delta LogWs/\delta$ AROM.FUSED[2] | $\delta LogWs/\delta$ AROM.FUSED[3] | $\delta LogWs/\delta$ AROM.FUSED[4p] |
|---|---|---|---|---|---|---|
| 1.0 | 252.31 | 12 | 8 | 1 | 1 | 2 |

To calculate the confidence intervals of estimated property values, say the 95% confidence intervals of the estimated $LogW_s$ value, the covariance matrix $COV(\boldsymbol{P}^*)$ given in Table 7 and

the local sensitivity $J(\boldsymbol{P}^*)$ given in Table 8 are substituted in eq. (9). For 95% confidence interval calculation, the t-distribution value corresponding to 0.05 /2 percentile (i.e. $\alpha_t$ /2 percentile) and with 4311 degrees of freedom (taken from Table 2) is obtained by solving eq. 8 for $t$ and this value is 1.9604. The predicted value of the $LogW_s$ is -2.64 (see Table 6). The calculated 95% confidence intervals of the estimated $LogW_s$ value is therefore,

$$\mathrm{LogW_s}^{pred}_{(1-0.05)} = \underbrace{\mathrm{LogW_s}^{pred}}_{-2.64} \pm \underbrace{\sqrt{\mathrm{diag}\left( J\left(\boldsymbol{P}^*\right) COV\left(\boldsymbol{P}^*\right) J\left(\boldsymbol{P}^*\right)^T \right)}}_{0.2134} . \underbrace{t\left(\nu, \alpha_t/2\right)}_{1.9604} = -2.64 \pm 0.41$$

It can be observed that the experimental value of the $LogW_s$ (-2.79) lies within the predicted confidence intervals indicating reliability of the developed model for estimating property values of $LogW_s$ and uncertainties of the estimated values. This, of course, can only be checked when experimental data is available. This is further illustrated in Figure 2 by plotting the experimental values of $LogW_s$ and the calculated 95% confidence intervals (shown as vertical bars) for the entire experimental data-set of $LogW_s$ used for the regression purpose. About 42% of the experimental values in the data-set (with 4681 data points) of $LogW_s$ falls within the confidence intervals calculated at 95% confidence level.

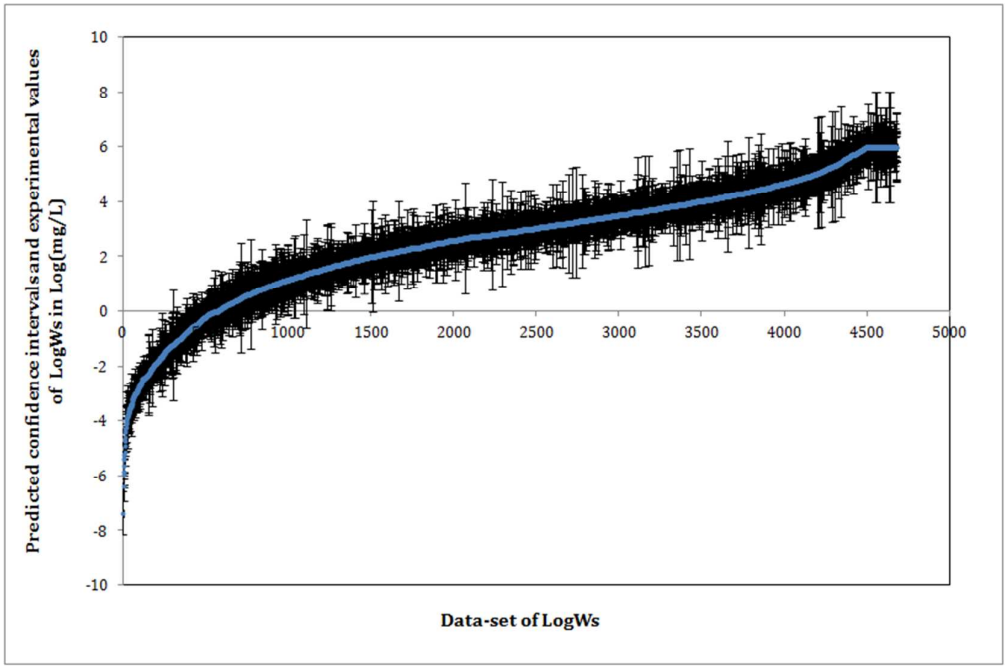



**Figure 2.** Experimental values of $LogW_s$ and the calculated 95% confidence intervals versus data-set of $LogW_s$

For the case where no experimental data is available, the calculated confidence intervals provide a measure of the likely prediction (that is, uncertainty) error of the predicted property value. We have considered here the calculation of confidence intervals of the estimated property values using models analysed by simultaneous regression method in order to simplify the illustration of the application of the developed property models, since there will be a single covariance matrix containing covariance of all the listed groups and universal parameters. The approach discussed in this section is the same for the case of property models analysed using the step-wise regression method. In the case of step-wise regression method, there will be a covariance matrix for each type of the groups, i.e., first-order, second-order and third-order and hence, quantification of uncertainty in the predicted property value is to be performed (using these covariance matrices) for each step (that is step1, step 2, and step 3) of property estimation.

## CONCLUSIONS

Property models for environment-related properties based on the $GC^+$ approach have been developed with the objective of providing reliable estimation of these properties together with the uncertainties of the estimated values for their use in the synthesis, design, and analysis of sustainable processes. The estimation of environment-related properties using these models requires only the molecular structure of the organic chemicals. Large experimental data-sets of environment-related properties taken from the database of US Environmental Protection Agency (EPA) and from the database of USEtox$^{TM}$ are used for the regression purpose in order to achieve good model performance and large application range of the property models. In total 22 environment-related properties of organic chemicals have been modelled and analysed. The use of the developed property models to estimate environment-related properties and the uncertainties of the estimated property values is illustrated through an

application example. The models for some of the properties analysed in this work have been implemented in ProPred, a property estimation toolbox of ICAS ® (Integrated Computer Aided System) software. The developed property models provide reliable estimates of environment-related properties needed to perform design and analysis of sustainable processes and allow one to evaluate the effect of uncertainties of estimated property values on the calculated potential impact that the processes would have on the environment giving useful insights into quality and reliability of the design of sustainable processes. Our current and future work is focused on quantification of the effect of uncertainties of estimated properties on the design of sustainable processes.

## ABBREVIATIONS

AAE average absolute error

$a_i$ contribution of atom of type-$i$

$A_i$ occurrence of atom of type-$i$

ARE average relative error [%]

$b$ adjustable parameter of Eq. (6)

$B$ (1/2, ν/2) beta function

$c$ adjustable parameter of Eq. (6)

CI atom connectivity index

$C_i$ contribution of first-order group of type-$i$

$COV(\boldsymbol{P}^*)$ covariance matrix

$d$ universal parameter of Eq. (6)

$D_j$ contribution of second-order group of type-$j$

$E_k$ contribution of third-order group of type-$k$

$f(X)$ function of property $X$

GC group-contribution

$GC^+$ group-contribution$^+$

$J(\boldsymbol{P}^*)$ local sensitivity of the model to variations in estimated model parameters

MG Marrero and Gani

$M_j$ occurrence of second-order group of type-$j$

*MSECV* mean squared error of cross-validation

N number of experimental data-points used in the regression

$N_i$ occurrence of first-order group of type-$i$

$O_k$ occurrence of third-order group of type-$k$

$\boldsymbol{P}$ model parameters

$\boldsymbol{P}^*$ estimated values of model parameters

$P_r(t, v)$ Students t-distribution function

$P_{rc}$ percentage of the experimental data-points [%]

$R$ universal gas constant [cc-bar/mol-K]

$R^2$ coefficient of determination

$S(\boldsymbol{P})$ cost function

SD standard deviation

*SSE* minimum sum of squared errors

$t(v, \alpha_t/2)$ t-distribution value corresponding to the $\alpha_t/2$ percentile

$X^{exp}$ experimental property value

$X^{pred}$ predicted property value

**Greek Symbols**

$^{v}\chi^{0}$ zeroth-order (atom) connectivity index

$^{v}\chi^{1}$ first-order (bond) connectivity index

$\nu$ degrees of freedom

## ACKNOWLEDGMENT

The authors acknowledge U.S. EPA, National Risk Management Research Laboratory, Cincinnati, Ohio for providing experimental data for this work. This research work is a part of the collaboration between Technical University of Denmark, Denmark and Alfa Laval Copenhagen A/S, Denmark. The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7-MC-ITN] under grant agreement no. 238013.

**SUPPORTING INFORMATION:** Following tables are included as a supporting information and is available free of charge via the Internet at http://pubs.acs.org, or at http://www.capec.kt.dtu.dk/documents/environment_related_properties/supporting_information.pdf.

Table S1. Performance of MG Method Based Property Models Analysed Using Simultaneous Regression Method

Table S2. MG Method Based Property Models Analysed Using Step-Wise Regression Method: First-Order Groups and their Contributions for the Environment-Related Properties

Table S3. MG Method Based Property Models Analysed Using Step-Wise Regression Method: Second-Order Groups and their Contributions for the Environment-Related Properties

Table S4. MG Method Based Property Models Analysed Using Step-Wise Regression Method: Third-Order Groups and their Contributions for the Environment-Related Properties

Table S5. MG Method Based Property Models Analysed Using Simultaneous Regression Method: First-Order Groups and their Contributions for the Environment-Related Properties

Table S6. MG Method Based Property Models Analysed Using Simultaneous Regression Method: Second-Order Groups and their Contributions for the Environment-Related Properties

Table S7. MG Method Based Property Models Analysed Using Simultaneous Regression Method: Third-Order Groups and their Contributions for the Environment-Related Properties

Table S8. CI Method Based Property Models: Atom Contributions and Model Constants for the Environment-Related Properties

## REFERENCES

(1) Carvalho, A.; Gani, R.; Matos, H. Design of sustainable processes: Systematic generation and evaluation of alternatives. *Comput.-Aided Chem. Eng*. **2006**, *21*, 817–822.

(2) Cabezas, H.; Bare, J. C.; Mallick, S. K. Pollution prevention with chemical process simulators: The generalized waste reduction (WAR) algorithm – full version. *Comput. Chem. Eng*. **1999***, 23*(4–5), 623–634.

(3) Young, D. M.; Cabezas, H. Designing sustainable processes with simulation: The waste reduction (WAR) algorithm. *Comput. Chem. Eng.* **1999**, *23*(10), 1477–1491.

(4) Young, D. M.; Sharp, R.; Cabezas, H. The waste reduction (WAR) algorithm: Environmental impacts, energy consumption, and engineering economics. *Waste Manage.* **2000**, *20*, 605–615.

(5) Jensen, N.; Coll, N.; Gani, R. An integrated computer-aided system for generation and evaluation of sustainable process alternatives. *Clean Technol. Environ. Policy*. **2003**, *5*, 209–225.

(6) Carvalho, A.; Gani, R.; Matos, H. Design of sustainable chemical processes: Systematic retrofit analysis generation and evaluation of alternatives. *Process Saf. Environ. Prot.* **2008**, *86*(5), 328–346.

(7) USEtox$^{TM}$ model. http://www.usetox.org/.

(8) Rosenbaum R. K.; Huijbregts M. A. J.; Henderson A. D.; Margni M.; McKone T. E.; van de Meent D.; Hauschild M. Z.; Shaked S.; Li D: S.; Gold L. S.; Jolliet O. USEtox human exposure and toxicity factors for comparative assessment of toxic emissions in life cycle analysis: sensitivity to key chemical properties. *Int. J. Life Cycle Assess.* **2011**, *16*, 710–727.

(9) Boethling, R. S.; Howard, P. H.; Meylan, W. M. Finding and estimating chemical property data for environmental assessment. *Environ. Toxicol. Chem.* **2004**, *23*, 2290-2308.

(10) Joback, K. K.; Reid, R. Estimation of pure component properties from group contribution. *Chem. Eng. Commun.* **1987**, *57*, 233-247.

(11) Lydersen, A. L. Estimation of critical properties of organic compounds. Report 3; University of Wisconsin, College of Engineering, Engineering Experimental Station: Madison, WI, 1955.

(12) Constantinou, L.; Gani, R. A new group contribution method for the estimation of properties of pure compounds. *AIChE Journal*. **2004**, *40*(10), 1697-1710.

(13) Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, *183-184*, 183-208.

(14) Gani, R.; Harper, P. M.; Hostrup, M. Automatic creation of missing groups through connectivity index for pure-component property prediction. *Ind. Eng. Chem. Res.* **2005**, *44*, 7262-7269.

(15) Martin, T. M.; Young, D.M. Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (*pimephales promelas*) using a group contribution method. *Chem. Res. Toxicol*. **2000**, *14*, 1378-1385.

(16) Casalegno, M.; Benfenati, E.; Sello, G. An automated group contribution method in predicting aquatic toxicity: the diatomic fragment approach. *Chem. Res. Toxicol*. **2005**, *18*, 740-746.

(17) Marrero, J.; Gani, R. A group contribution based estimation of octanol-water partition coefficient and aqueous solubility. *Ind. Eng. Chem. Res.* **2002**, *41*, 6623-6633.

(18) Klopman, G.; Zhu, H. Estimation of the aqueous solubility of organic molecules by the group contribution approach. *J. Chem. Inf. Model.* **2001**, *41*, 439-445.

(19) Kühne, R.; Ebert, R.-U.; Kleint, F.; Scmidt, G.; Schüürmann, G. Group contribution methods to estimate water solubility of organic chemicals. *Chemosphere*. **1995**, *30*, 2061-2077.

(20) Martin T.; Harten, P.; Venkatapathy, R.; Das, S.; Young, D. A Hierarchical Clustering Methodology for the Estimation of Toxicity. *Toxicol. Mech. Methods*. **2008**, *18*, 251-266.

(21) U.S. Environmental Protection Agency (2012a). Estimation Program Interface Suite (EPISUITE). http://www.epa.gov/oppt/exposure/pubs/episuite.htm.

(22) U.S. Environmental Protection Agency (2012b). Toxicity Estimation Software Tool (T.E.S.T.). http://www.epa.gov/nrmrl/std/qsar/qsar.html.

(23) Isitutio Mario Negri (2012). VEGA. http://www.vega-qsar.eu/contributors.html.

(24) Hukkerikar, A. S.; Sarup, B.; Ten Kate, A.; Abildskov, J.; Sin, G.; Gani, R. Group contribution$^{+}$ ($GC^{+}$) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilib.* **2012,** *321*, 25-43.

(25) Seber, G.; Wild, C. Nonlinear regression; Wiley: New York, 1989.

(26) Sin, G.; Meyer, A. S.; Gernaey, K.V. Assessing reliability of cellulose hydrolysis models to support biofuel process design – identifiability and uncertainty analysis. *Comput. Chem. Eng.* **2010**, *34*, 1385–1392.

(27) Abramowitz, M.; Stegun, I. A. Handbook of mathematical functions with formulas, graphs, and mathematical tables; New York: Dover, 1972.

(28) Mevik, B-H.; Cederkvist, H. R. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *J. Chemom*. **2004**, *18*, 422-429.

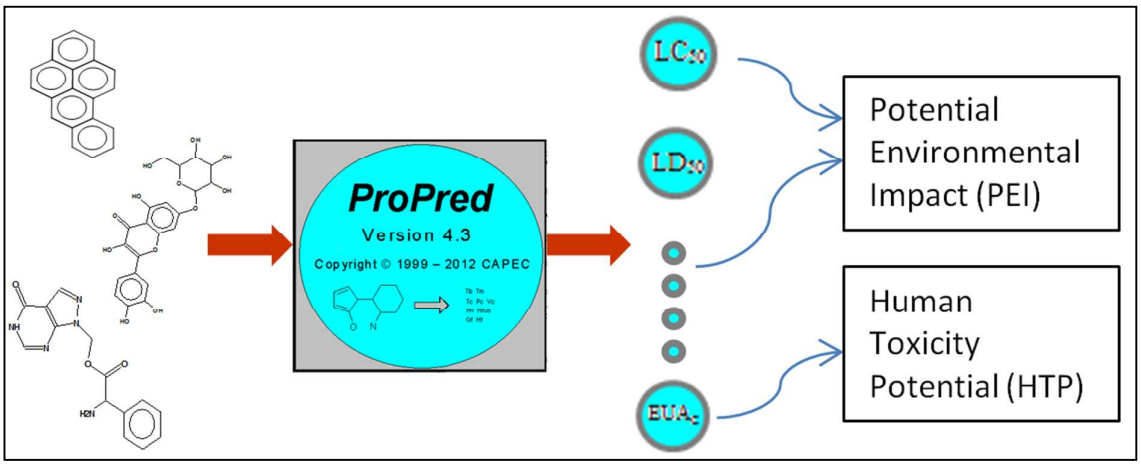
(29) Integrated Computer Aided System (ICAS), version 15.0; Department of Chemical Engineering, Technical University of Denmark, Lyngby, Denmark, 2012.

**FOR TABLE OF CONTENTS USE ONLY**

**TITLE**: Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution$^{+}$ (GC$^{+}$) Property Models and Uncertainty Analysis

**AUTHORS**: Amol Shivajirao Hukkerikar, Sawitree Kalakul, Bent Sarup, Douglas M. Young, Gürkan Sin, and Rafiqul Gani.

**TABLE OF CONTENTS (TOC) GRAPHIC**



## Figures and TABLE OF CONTENTS (TOC) GRAPHIC for the manuscript

**Figure 1.** Plot of primary alcohols versus their experimental values of $-\text{Log}LC_{50}$(FM)

**Figure 2.** Experimental values of $\text{Log}W_s$ and the calculated 95% confidence intervals versus data-set of $\text{Log}W_s$

**TABLE OF CONTENTS (TOC) GRAPHIC**

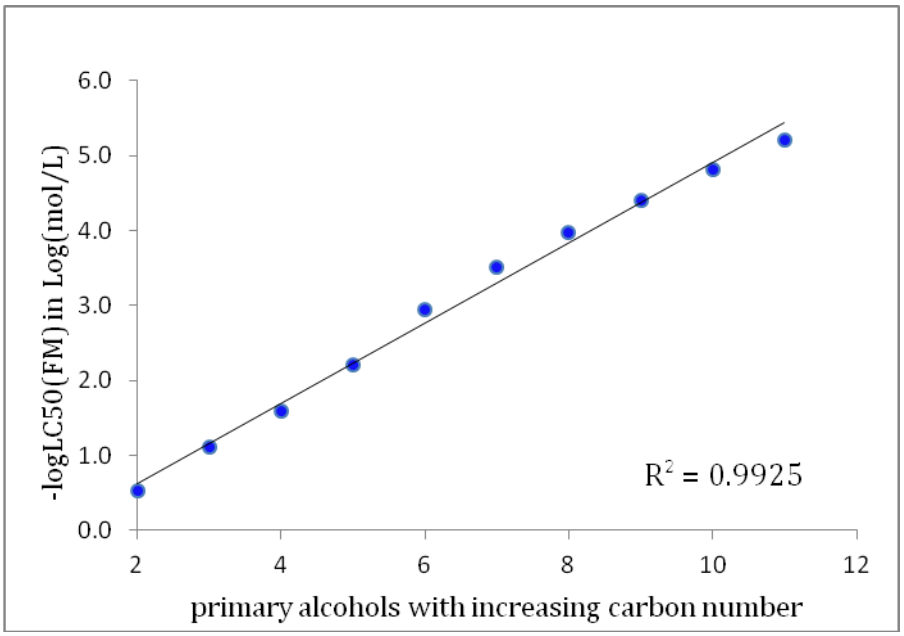



**Figure 1.** Plot of primary alcohols versus their experimental values of $-\text{LogLC}_{50}$(FM)

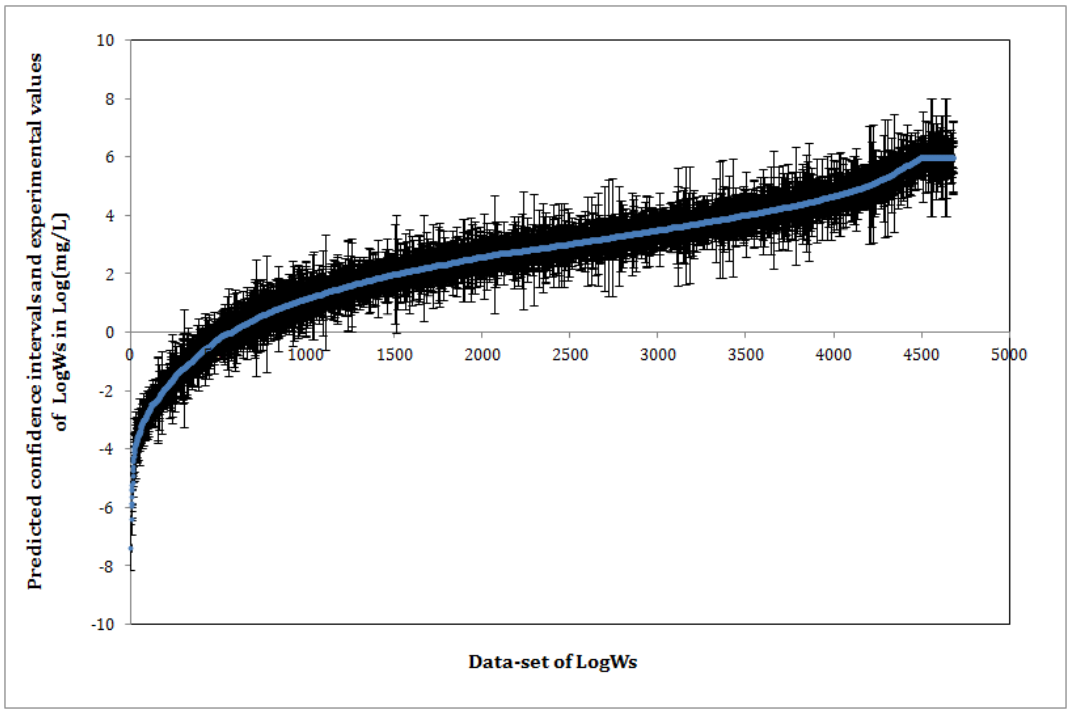
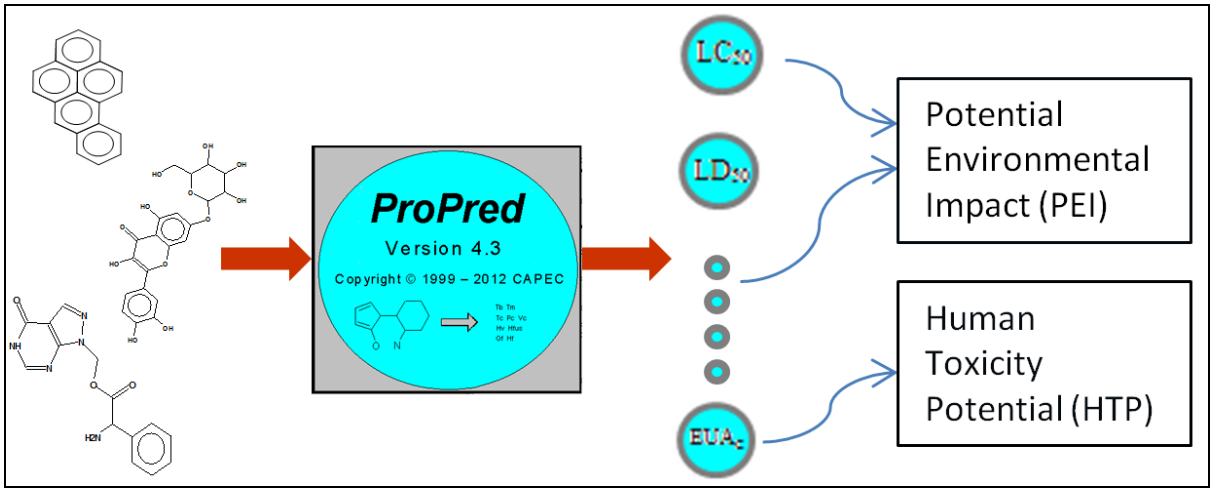



**Figure 2.** Experimental values of $LogW_s$ and the calculated 95% confidence intervals versus data-set of $LogW_s$

## TABLE OF CONTENTS (TOC) GRAPHIC



**SUPPORTING INFORMATION**

# Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution$^{+}$ (GC$^{+}$) Property Models and Uncertainty Analysis

*Amol Shivajirao Hukkerikar* [a], *Sawitree Kalakul* [b], *Bent Sarup* [c], *Douglas M. Young* [d], *Gürkan Sin* [a], *and Rafiqul Gani* [a,*]

[a] Computer Aided Process-Product Engineering Center (CAPEC), Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark.

[b] The Petroleum and Petrochemical College, Chulalongkorn University, Bangkok 10330, Thailand.

[c] Vegetable Oil Technology Business Unit, Alfa Laval Copenhagen A/S, Maskinvej 5, DK-2860, Søborg, Denmark.

[d] U.S. EPA, National Risk Management Research Laboratory, Cincinnati, Ohio 45268.

[*]Corresponding author: E-mail address: rag@kt.dtu.dk (R. Gani); Phone no: +45-45252882; Fax no: +45-45932906.

Following tables are included as supporting information.

Table S1. Performance of MG Method Based Property Models Analysed Using Simultaneous Regression Method

Table S2. MG Method Based Property Models Analysed Using Step-Wise Regression Method: First-Order Groups and their Contributions for the Environment-Related Properties

Table S3. MG Method Based Property Models Analysed Using Step-Wise Regression Method: Second-Order Groups and their Contributions for the Environment-Related Properties

Table S4. MG Method Based Property Models Analysed Using Step-Wise Regression Method: Third-Order Groups and their Contributions for the Environment-Related Properties

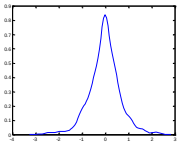
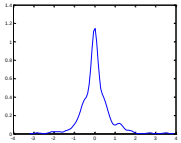
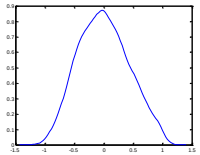
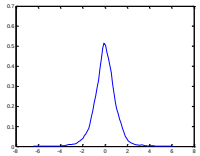
Table S5. MG Method Based Property Models Analysed Using Simultaneous Regression Method: First-Order Groups and their Contributions for the Environment-Related Properties

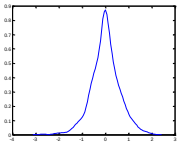
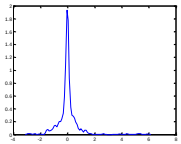
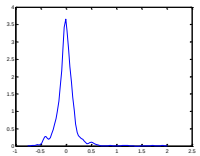
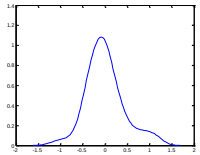
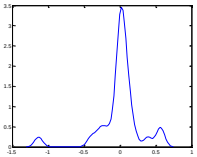
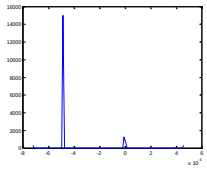
Table S6. MG Method Based Property Models Analysed Using Simultaneous Regression Method: Second-Order Groups and their Contributions for the Environment-Related Properties

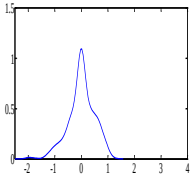
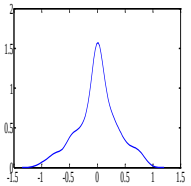
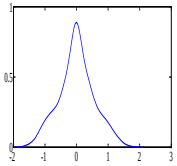
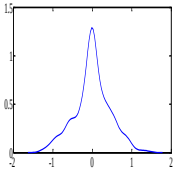
Table S7. MG Method Based Property Models Analysed Using Simultaneous Regression Method: Third-Order Groups and their Contributions for the Environment-Related Properties

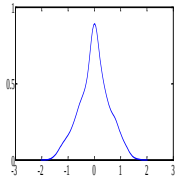
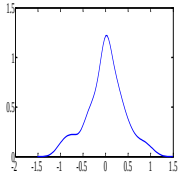
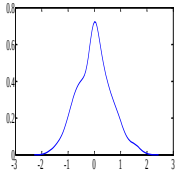
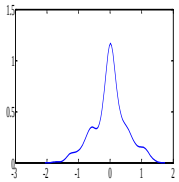
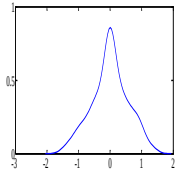
Table S8. CI Method Based Property Models: Atom Contributions and Model Constants for the Environment-Related Properties

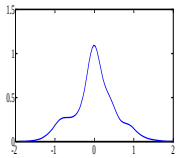
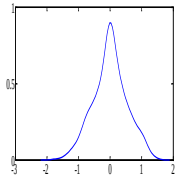
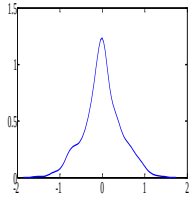
**Table S1. Performance of MG Method Based Property Models Analysed Using Simultaneous Regression Method**

| sl. no. | property | L.H.S. of MG method based property prediction model | MG group-contribution model $f(X)=\sum_i N_i C_i + \sum_j M_j D_j + \sum_k E_k O_k$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $f(X)$ | N | $\nu$ | $R^2$ | residual distribution plot | $P_{rc}$ (± 1% ) | $P_{rc}$ (± 5% ) | $P_{rc}$ (± 10%) | SD | AAE | ARE [a] |
| 1 | fathead minnow 96-hr $LC_{50}$ ($LC_{50}$ (FM)) in mol/lit | $-\mathrm{LogLC_{50}(FM)} + \mathrm{FM_0}$ | 809 | 541 | 0.81 |  | 9.64 | 33.00 | 57.11 | 0.63 | 0.45 | 20.14 |
| 2 | *daphnia magna* 48-hr $LC_{50}$ ($LC_{50}$ (DM)) in mol/lit | $-\mathrm{LogLC_{50}(DM)} + \mathrm{DM_0}$ | 320 | 124 | 0.87 |  | 28.75 | 48.13 | 67.50 | 0.61 | 0.38 | 11.19 |
| 3 | oral Rat $LD_{50}$ ($LD_{50}$) in mol/kg | $-\mathrm{LogLD50} - \mathrm{A_{LD50}} - \mathrm{B_{LD50}MW}$ | 5995 | 5617 | 0.74 |  | 1.87 | 7.96 | 14.18 | 0.42 | 0.35 | 16.09 |
| 4 | aqueous solubility ($LogW_S$) in gm/lit | $\log(W_S) - \mathrm{A_{W_S}} - \mathrm{B_{W_s}MW}$ | 4681 | 4311 | 0.79 |  | 3.23 | 14.80 | 28.97 | 0.97 | 0.71 | ---- |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | bioconcentration factor (BCF) | LogBCF | 662 | 423 | 0.80 |  | 9.37 | 20.85 | 30.36 | 0.60 | 0.44 | ---- | |
| 6 | permissible Exposure Limit (PEL) ) in mol/lit | -LogPEL | 425 | 239 | 0.78 |  | 23.53 | 48.47 | 65.88 | 0.72 | 0.38 | 11.02 | |
| 7 | photochemical Oxidation Potential (PCO) | -LogPCO | 639 | 488 | 0.86 |  | 5.95 | 18.31 | 28.64 | 0.20 | 0.12 | 6.60 | |
| 8 | global Warming Potential (GWP) | LogGWP | 51 | 31 | 0.87 |  | 15.69 | 37.25 | 56.86 | 0.41 | 0.29 | 11.57 | |
| 9 | ozone Depletion Potential (ODP) | LogODP | 28 | 12 | 0.89 |  | 17.86 | 21.43 | 28.5 | 0.30 | 0.16 | ---- | |
| 10 | acidification Potential (ODP) | LogAP | 10 | 1 | 1.0 |  | 100.0 | -- | -- | 3.4E-04 | 2.1E-4 | ---- | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | emission to urban air ($EUA_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(\mathrm{EUA}_C)+A_{\mathrm{EUA}_C}$ | 456 | 214 | 0.84 |  | 23.03 | 48.90 | 75.22 | | 0.51 | 0.36 | 7.42 |
| 12 | emission to urban air ($EUA_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(\mathrm{EUA}_{NC})+A_{\mathrm{EUA}_{NC}}$ | 341 | 128 | 0.88 |  | 27.57 | 60.70 | 84.46 | | 0.36 | 0.26 | 4.87 |
| 13 | emission to continental rural air ($ERA_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(\mathrm{ERA}_C)+A_{\mathrm{ERA}_C}$ | 470 | 229 | 0.83 |  | 20.43 | 45.53 | 71.91 | | 0.56 | 0.42 | 7.18 |
| 14 | emission to continental rural air ($ERA_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(\mathrm{ERA}_{NC})+A_{\mathrm{ERA}_{NC}}$ | 349 | 134 | 0.87 |  | 26.36 | 55.30 | 80.52 | | 0.45 | 0.32 | 5.50 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | emission to continental fresh water ($EFW_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(EFW_C)+A_{EFW_C}$ | 472 | 230 | 0.83 |  | 20.34 | 41.95 | 68.43 | 0.55 | 0.41 | 8.87 |
| 16 | emission to continental fresh water ($EFW_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(EFW_{NC})+A_{EFW_{NC}}$ | 345 | 131 | 0.89 |  | 23.19 | 52.75 | 77.97 | 0.42 | 0.31 | 6.33 |
| 17 | emission to continental sea water ($ESW_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ESW_C)+A_{ESW_C}$ | 477 | 235 | 0.87 |  | 22.22 | 46.96 | 76.10 | 0.63 | 0.48 | 6.72 |
| 18 | emission to continental sea water ($ESW_{NC}$) in cases/kg emitted (non-carcinogenic) | $-\mathrm{Log}(ESW_{NC})+A_{ESW_{NC}}$ | 360 | 146 | 0.91 |  | 23.33 | 58.06 | 80.56 | 0.52 | 0.38 | 6.06 |
| 19 | emission to continental natural soil ($ENS_C$) in cases/kg emitted (carcinogenic) | $-\mathrm{Log}(ENS_C)+A_{ENS_C}+B_{ENS_C}\times M$ | 460 | 219 | 0.84 |  | 24.13 | 50.00 | 76.52 | 0.52 | 0.38 | 6.22 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | emission to continental natural soil ($ENS_{NC}$) in cases/kg emitted (non-carcinogenic) | $-Log(ENS_{NC})+A_{ENS_{NC}}$ | 362 | 148 | 0.85 |  | 23.48 | 55.25 | 77.90 | 0.51 | 0.37 | 5.91 |
| 21 | emission to continental agri-cultural soil ($EAS_C$) in cases/kg emitted (carcinogenic) | $-Log(EAS_C)+A_{EAS_C}$ | 470 | 228 | 0.84 |  | 22.34 | 47.23 | 73.83 | 0.53 | 0.40 | 6.91 |
| 22 | emission to continental agri-cultural soil ($EAS_{NC}$) in cases/kg emitted (non-carcinogenic) | $-Log(EAS_{NC})+A_{EAS_{NC}}$ | 352 | 138 | 0.87 |  | 23.86 | 56.53 | 82.39 | 0.44 | 0.32 | 5.30 |

[a] ARE is not defined for LogWs, BCF, ODP and AP since these properties have both positive and negative values.

**Table S2. MG Method Based Property Models Analysed Using Step-Wise Regression Method: First-Order Groups and their Contributions [a] for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), PCO, GWP, ODP, and AP.**

| | Group | $LC50(FM)_{1i}$ | $LC50(DM)_{1i}$ | $LD50_{1i}$ | $LogWs_{1i}$ | $BCF_{1i}$ | $PEL_{1i}$ | $PCO_{1i}$ | $GWP_{1i}$ | $ODP_{1i}$ | $AP_{1i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $CH_3$ | 0.0972 | -0.0386 | -0.0742 | -5.2494 | 0.6657 | 0.7723 | 0.1227 | 0.3880 | -0.9453 | -0.1290 |
| 2 | $CH_2$ | 0.2885 | 0.1710 | 0.0223 | -5.0706 | 0.0948 | 0.0727 | 0.0463 | -1.0699 | **** | **** |
| 3 | CH | 0.2441 | -0.1654 | 0.1335 | -4.8948 | -0.3921 | -0.6557 | -0.0790 | **** | **** | **** |
| 4 | C | -0.3822 | 0.4640 | 0.2641 | -4.6277 | -0.9137 | -1.3404 | -0.0434 | **** | **** | **** |
| 5 | $CH_2$=CH | 1.0340 | 0.1698 | 0.1087 | -9.6240 | 0.7712 | 2.2638 | -0.2572 | **** | **** | **** |
| 6 | CH=CH | 0.3890 | -0.2512 | 0.0977 | -9.5597 | **** | 0.7658 | -0.5513 | **** | **** | **** |
| 7 | $CH_2$=C | 0.5436 | 0.5657 | 0.1358 | -9.4175 | -0.1358 | 0.1652 | -0.3932 | **** | **** | **** |
| 8 | CH=C | 0.5902 | 0.6270 | 0.2376 | -9.4095 | -0.0639 | -0.7468 | -0.6306 | **** | **** | **** |
| 9 | C=C | 0.8302 | 0.2129 | 0.3451 | -9.0081 | -0.1808 | -1.2669 | -0.6010 | **** | **** | **** |
| 10 | $CH_2$=C=CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | $CH_2$=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 12 | C=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | CH≡C | 0.7491 | **** | 0.0609 | -9.0073 | **** | 3.0138 | -0.3025 | **** | **** | **** |
| 14 | C≡C | 1.6682 | -0.2044 | 0.3876 | -8.6808 | **** | **** | -0.9074 | **** | **** | **** |
| 15 | aCH | 0.1530 | 0.0743 | 0.0230 | -4.6135 | 0.2561 | 0.4045 | 0.0137 | **** | **** | **** |
| 16 | aC fused with aromatic ring | 0.1782 | 0.2851 | 0.0398 | -4.7032 | -0.0341 | 0.2664 | 0.0053 | **** | **** | **** |
| 17 | aC fused with non-aromatic ring | 0.6141 | 0.4121 | 0.1363 | -4.5533 | 0.1641 | 0.3003 | -0.0341 | **** | **** | **** |
| 18 | aC except as above | 0.3150 | 0.3864 | 0.0501 | -4.5067 | -0.0071 | -0.0178 | 0.8000 | **** | **** | **** |
| 19 | aN in aromatic ring | -0.1501 | -0.2125 | 0.0755 | -4.4745 | 0.0130 | 1.4732 | **** | **** | **** | **** |
| 20 | aC-$CH_3$ | 0.4050 | 0.4345 | 0.0699 | -9.6580 | 0.4333 | 0.5522 | -0.1132 | **** | **** | **** |
| 21 | aC-$CH_2$ | 0.3350 | 0.2087 | 0.0821 | -9.3278 | -0.1904 | -0.3934 | -0.2018 | **** | **** | **** |
| 22 | aC-CH | 0.7893 | 0.9717 | 0.4545 | -9.0233 | -0.9151 | -1.1787 | -0.2998 | **** | **** | **** |
| 23 | aC-C | 0.5639 | 1.0938 | 0.4264 | -8.7328 | -0.9509 | -0.9892 | -0.3527 | **** | **** | **** |
| 24 | aC-CH=$CH_2$ | 0.8110 | 0.0652 | 0.1599 | -14.0398 | 0.8698 | 0.5330 | 0.3822 | **** | **** | **** |
| 25 | aC-CH=CH | 0.5427 | **** | 0.0900 | -14.4421 | 0.1297 | **** | 0.8379 | **** | **** | **** |
| 26 | aC-C=$CH_2$ | **** | **** | -0.6814 | -14.2994 | 0.3000 | -0.4032 | 0.1949 | **** | **** | **** |
| 27 | aC-C≡CH | **** | **** | **** | -13.6670 | **** | **** | **** | **** | **** | **** |
| 28 | aC-C≡C | **** | **** | -0.3680 | -14.4769 | **** | **** | **** | **** | **** | **** |
| 29 | OH | -0.6115 | -0.8815 | -0.1955 | -5.1862 | -0.0340 | 1.3612 | 0.0359 | **** | **** | **** |
| 30 | aC-OH | 0.2670 | 0.2670 | 0.0705 | -9.6989 | -0.2700 | 1.2393 | 0.3113 | **** | -1.5842 | -0.0769 |
| 31 | COOH | -0.1104 | -0.3982 | 0.0320 | -14.8398 | -0.8830 | 2.3281 | -0.0379 | **** | **** | **** |
| 32 | aC-COOH | 0.0172 | -1.6411 | 0.0391 | -19.4187 | -0.9232 | 1.1988 | **** | **** | **** | **** |
| 33 | $CH_3CO$ | 0.0835 | -1.4018 | -0.0172 | -14.2841 | 0.6654 | 1.4016 | 0.1409 | **** | **** | **** |
| 34 | $CH_2CO$ | -0.5508 | **** | 0.1931 | -14.6443 | -1.4842 | 1.2601 | -0.0515 | **** | **** | **** |
| 35 | CHCO | **** | **** | 0.4130 | -15.1578 | **** | **** | 0.0505 | **** | **** | **** |
| 36 | CCO | **** | **** | 1.1655 | -14.0177 | **** | 0.1081 | **** | **** | **** | **** |
| 37 | aC-CO | 0.2699 | 0.4353 | 0.2190 | -14.1140 | -0.9893 | 0.9671 | **** | **** | **** | **** |
| 38 | CHO | 0.6008 | 0.7531 | -0.1338 | -9.3615 | -0.3560 | 2.3662 | -0.1855 | **** | **** | **** |
| 39 | aC-CHO | 0.8678 | 0.9419 | -0.0626 | -14.4954 | -1.3452 | **** | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | $CH_3COO$ | 0.4393 | -0.6598 | -0.1734 | -19.9009 | 0.3988 | 1.2544 | 0.3858 | **** | **** | **** |
| 41 | $CH_2COO$ | 0.1823 | 0.4007 | -0.0357 | -19.7361 | 0.2378 | 1.6798 | 0.3021 | **** | **** | **** |
| 42 | CHCOO | -0.0242 | **** | 0.1329 | -19.3313 | **** | **** | 0.1068 | **** | **** | **** |
| 43 | CCOO | -1.9110 | **** | 0.3242 | -19.3558 | **** | **** | **** | **** | **** | **** |
| 44 | HCOO | **** | **** | -0.0612 | -15.4178 | **** | 1.5780 | 0.8282 | **** | **** | **** |
| 45 | aC-COO | 0.4382 | 0.0355 | -0.0431 | -19.3244 | -1.1633 | 0.5326 | -0.0378 | **** | **** | **** |
| 46 | aC-OOCH | **** | **** | -0.2398 | **** | **** | **** | **** | **** | **** | **** |
| 47 | aC-OOC | 0.9868 | **** | 0.4845 | -19.5174 | -2.9904 | 0.9678 | **** | **** | **** | **** |
| 48 | COO except as above | 0.4138 | 1.3113 | 0.0657 | -14.9256 | -0.5196 | 0.4342 | -0.2744 | **** | **** | **** |
| 49 | $CH_3O$ | -0.5209 | 0.7930 | -0.0259 | -10.1615 | 0.2530 | 2.1251 | 0.1499 | 0.1245 | **** | **** |
| 50 | $CH_2O$ | -0.2160 | 0.2341 | 0.0974 | -9.9546 | -0.1060 | 0.9276 | -0.1226 | **** | **** | **** |
| 51 | CH-O | -0.8189 | **** | 0.4987 | -9.8757 | -0.3114 | -0.7462 | -0.2064 | -1.8521 | **** | **** |
| 52 | C-O | **** | **** | 0.2275 | -9.2509 | -0.7149 | **** | **** | **** | **** | **** |
| 53 | aC-O | 0.0896 | 0.7875 | 0.1839 | -9.6929 | -0.3264 | 0.4621 | -0.4603 | **** | **** | **** |
| 54 | $CH_2NH_2$ | -0.0246 | 0.1531 | 0.0450 | -9.5161 | 0.2885 | 1.9265 | -0.3290 | **** | **** | **** |
| 55 | $CHNH_2$ | -0.0983 | 0.9878 | 0.3764 | -10.1966 | **** | 2.1480 | -0.5376 | **** | **** | **** |
| 56 | $CNH_2$ | 0.5803 | **** | 0.4151 | -8.7982 | **** | **** | **** | **** | **** | **** |
| 57 | $CH_3NH$ | -0.4593 | 0.1983 | -0.0593 | -9.5708 | **** | 2.8505 | 0.0046 | **** | **** | **** |
| 58 | $CH_2NH$ | 0.0524 | 0.0553 | 0.2571 | -9.2845 | -0.3334 | 1.2126 | -0.0060 | **** | **** | **** |
| 59 | CHNH | **** | -0.3115 | 0.2506 | -8.8957 | **** | 1.2708 | **** | **** | **** | **** |
| 60 | $CH_3N$ | 0.1273 | -0.9599 | 0.3338 | -9.5015 | -0.4163 | 1.1981 | -0.4139 | **** | **** | **** |
| 61 | $CH_2N$ | -0.3121 | -1.4049 | 0.4337 | -9.3647 | -1.4607 | 0.2724 | -0.3483 | **** | **** | **** |
| 62 | aC-$NH_2$ | 0.0721 | 1.0918 | 0.1189 | -9.8417 | -0.3313 | 2.0982 | **** | **** | **** | **** |
| 63 | aC-NH | 0.3195 | 0.4657 | 0.1796 | -9.7080 | -0.4387 | 0.7464 | **** | **** | **** | **** |
| 64 | aC-N | 0.0691 | 0.5173 | 0.2682 | -9.5308 | -0.6948 | -0.7120 | **** | **** | **** | **** |
| 65 | $NH_2$ except as above | 0.1130 | 0.2107 | 0.0130 | -5.2748 | -0.1643 | 2.0449 | -0.4319 | **** | **** | **** |
| 66 | CH=N | 2.4719 | 0.0832 | 0.1706 | -9.5889 | 0.1155 | **** | **** | **** | **** | **** |
| 67 | C=N | 0.9533 | 1.1620 | 0.2576 | -8.8434 | -1.0204 | **** | **** | **** | **** | **** |
| 68 | $CH_2CN$ | -0.0683 | **** | -0.0168 | -13.5365 | 0.3979 | 2.0501 | **** | **** | **** | **** |
| 69 | CHCN | 2.0149 | **** | 1.0312 | -13.2307 | -0.4915 | 1.9526 | **** | **** | **** | **** |
| 70 | CCN | 1.0777 | **** | 0.7798 | -12.9126 | **** | 0.7938 | **** | **** | **** | **** |
| 71 | aC-CN | 0.0788 | 0.6363 | 0.1041 | -13.5595 | -0.3186 | **** | **** | **** | **** | **** |
| 72 | CN except as above | 0.5895 | 0.2307 | 0.3921 | -9.2357 | 1.2018 | 2.5607 | 0.4704 | **** | **** | **** |
| 73 | $CH_2NCO$ | **** | **** | 0.0453 | **** | **** | 3.1956 | **** | **** | **** | **** |
| 74 | CHNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 75 | CNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 76 | aC-NCO | **** | **** | -0.1489 | **** | **** | 2.0353 | -0.0504 | **** | **** | **** |
| 77 | $CH2NO_2$ | **** | **** | 0.1374 | -20.3294 | **** | 1.8813 | 1.0797 | **** | **** | **** |
| 78 | $CHNO_2$ | **** | **** | 0.6535 | -20.6143 | -1.0104 | 1.4511 | **** | **** | **** | **** |
| 79 | $CNO_2$ | 1.4039 | **** | 1.1115 | -18.8805 | **** | 0.1482 | **** | **** | **** | **** |
| 80 | aC-$NO_2$ | 0.7014 | 0.2756 | 0.2627 | -20.3307 | -0.0368 | 1.5088 | 1.3371 | **** | **** | **** |
| 81 | $NO_2$ except as above | 1.0513 | -1.3024 | -0.0320 | -16.4594 | **** | 1.4137 | **** | **** | **** | -0.0775 |
| 82 | ONO | **** | **** | 0.8495 | **** | **** | **** | **** | **** | **** | **** |
| 83 | $ONO_2$ | -0.0796 | -0.3048 | 0.2883 | -21.5057 | **** | 2.0086 | **** | **** | **** | -0.0775 |
| 84 | $HCON(CH_2)_2$ | -0.2502 | **** | **** | -22.5375 | **** | **** | **** | **** | **** | **** |
| 85 | $HCONHCH_2$ | **** | **** | **** | -19.5759 | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | $CONH_2$ | -0.1815 | 1.2951 | 0.1678 | -14.6677 | **** | 3.1108 | **** | **** | **** | **** |
| 87 | $CONHCH_3$ | 1.2999 | -0.2366 | 1.0423 | -19.5825 | -0.6677 | **** | **** | **** | **** | **** |
| 88 | $CONHCH_2$ | 0.9036 | 3.7503 | 0.1446 | -19.3632 | -1.7679 | **** | **** | **** | **** | **** |
| 89 | $CON(CH_3)_2$ | -0.9938 | **** | 0.7419 | -23.9555 | 1.7095 | 2.6238 | **** | **** | **** | **** |
| 90 | $CONCH_3CH_2$ | **** | **** | **** | -23.7463 | **** | **** | **** | **** | **** | **** |
| 91 | $CON(CH_2)_2$ | -0.5947 | -0.5260 | 0.3286 | -23.5108 | -0.6649 | **** | **** | **** | **** | **** |
| 92 | CONHCO | **** | **** | 0.5830 | -24.4140 | **** | **** | **** | **** | **** | **** |
| 93 | CONCO | 1.3874 | **** | 0.4392 | -23.7170 | **** | **** | **** | **** | **** | **** |
| 94 | $aC\text{-}CONH_2$ | -0.3693 | -1.3986 | 0.2374 | -19.1965 | **** | **** | **** | **** | **** | **** |
| 95 | aC-NH(CO)H | -0.0863 | **** | 0.2940 | -18.3534 | **** | **** | **** | **** | **** | **** |
| 96 | aC-N(CO)H | -0.3164 | **** | 0.4427 | **** | **** | **** | **** | **** | **** | **** |
| 97 | aC-CONH | 1.2356 | -2.8641 | 0.0444 | -18.6958 | -1.8423 | **** | **** | **** | **** | **** |
| 98 | aC-NHCO | -0.1143 | 0.0285 | 0.0436 | -18.9709 | -1.2286 | **** | **** | **** | **** | **** |
| 99 | aC-(N)CO | 0.1407 | -0.0767 | 0.2094 | -18.1536 | -2.0069 | **** | **** | **** | **** | **** |
| 100 | NHCONH | **** | **** | 0.1069 | -20.4704 | -3.0761 | **** | **** | **** | **** | **** |
| 101 | $NH_2CONH$ | **** | **** | -0.0066 | -20.2551 | **** | **** | **** | **** | **** | **** |
| 102 | $NH_2CON$ | **** | **** | 0.7208 | -19.6286 | **** | **** | **** | **** | **** | **** |
| 103 | NHCON | -0.6989 | **** | 0.5454 | -19.0554 | **** | **** | **** | **** | **** | **** |
| 104 | NCON | **** | -1.2393 | 0.4743 | -17.1827 | **** | **** | **** | **** | **** | **** |
| 105 | $aC\text{-}NHCONH_2$ | **** | **** | 0.0291 | -24.4105 | **** | **** | **** | **** | **** | **** |
| 106 | aC-NHCONH | -0.5009 | 2.1330 | 0.0802 | -24.9922 | -1.4420 | **** | **** | **** | **** | **** |
| 107 | NHCO except as above | 0.6549 | -0.9079 | 0.1199 | -14.2024 | 0.6908 | **** | **** | **** | **** | **** |
| 108 | $CH_2Cl$ | 0.6181 | -0.1689 | 0.3896 | -17.3359 | 0.4043 | 2.3309 | 0.3730 | -0.2757 | -1.4515 | **** |
| 109 | CHCl | 0.1069 | 0.5411 | 0.2054 | -17.1019 | 0.3504 | -0.9263 | 0.6117 | **** | -1.3742 | **** |
| 110 | CCl | **** | **** | 0.2454 | -17.0334 | **** | **** | **** | **** | **** | **** |
| 111 | $CHCl_2$ | 0.5662 | 0.3245 | 0.2650 | -29.1152 | 0.4145 | 1.7348 | 1.5265 | -0.0233 | -1.7873 | **** |
| 112 | $CCl_2$ | 0.7493 | **** | -0.0580 | -28.3154 | 0.8521 | **** | **** | **** | **** | **** |
| 113 | $CCl_3$ | 1.3517 | 1.2019 | 0.3039 | -41.6097 | 0.9977 | 1.8541 | 2.7667 | 1.8173 | -0.0399 | -0.0132 |
| 114 | $CH_2F$ | **** | **** | 2.1620 | -10.7960 | **** | **** | 1.9489 | 0.8584 | **** | **** |
| 115 | CHF | **** | **** | **** | -9.2501 | **** | **** | **** | -0.6900 | **** | **** |
| 116 | CF | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 117 | $CHF_2$ | 1.0749 | **** | 0.3702 | -17.5991 | **** | **** | 1.4794 | 1.6870 | -0.5628 | **** |
| 118 | $CF_2$ | 0.1261 | **** | 0.1139 | -18.6216 | 0.6323 | -0.6817 | **** | -0.0026 | -0.2492 | **** |
| 119 | $CF_3$ | 0.6351 | -0.4425 | 1.3466 | -24.1014 | 0.3278 | 1.3676 | 0.5828 | 2.1289 | 0.0883 | **** |
| 120 | $CCl_2F$ | **** | **** | -0.2211 | -36.0704 | 0.8196 | 0.5580 | 2.4794 | 2.1575 | 0.0131 | 0.1546 |
| 121 | HCClF | **** | **** | -0.4169 | -23.5774 | **** | 0.0865 | 2.0994 | 0.9439 | -1.4911 | **** |
| 122 | $CClF_2$ | **** | **** | 0.1143 | -30.0684 | 0.8033 | 0.4950 | 2.4794 | 2.0786 | -0.0409 | **** |
| 123 | aC-Cl | 0.6323 | 0.4909 | 0.1991 | -17.1355 | 0.6407 | 0.9796 | 0.6083 | **** | -0.2832 | -0.0290 |
| 124 | aC-F | 0.3392 | 0.6346 | 0.2812 | -11.0066 | 0.5743 | **** | 0.3160 | **** | **** | **** |
| 125 | aC-I | 0.9439 | **** | 0.2841 | -47.8855 | 0.7695 | **** | **** | **** | **** | **** |
| 126 | aC-Br | 0.7252 | 0.8113 | 0.4920 | -32.2539 | 0.3579 | **** | **** | **** | **** | **** |
| 127 | -I except as above | 0.9066 | -0.2044 | 0.4522 | -43.7256 | **** | 2.9327 | **** | -2.1289 | **** | **** |
| 128 | -Br except as above | 0.6769 | 0.4380 | 0.3970 | -27.6030 | 0.5466 | 1.5345 | 0.9433 | 1.0645 | 0.4320 | **** |
| 129 | -F except as above | 0.0814 | 0.1183 | -0.0353 | -6.6241 | 0.5813 | 1.2998 | 0.4224 | 1.9617 | -0.0074 | **** |
| 130 | -Cl except as above | 0.3776 | 0.1975 | 0.0133 | -12.4658 | 0.5005 | 1.3093 | 0.5640 | 1.2757 | -0.0056 | -0.0682 |
| 131 | CHNOH | 0.7400 | **** | 0.2486 | -15.5688 | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 132 | CNOH | -0.4679 | **** | 0.1320 | -14.8149 | -1.2471 | **** | **** | **** | **** | **** |
| 133 | aC-CHNOH | 1.5961 | **** | 0.9452 | **** | **** | **** | **** | **** | **** | **** |
| 134 | $OCH_2CH_2OH$ | -0.7009 | -1.2305 | -0.1547 | -19.6115 | -0.2796 | 1.7163 | 0.0024 | **** | **** | **** |
| 135 | $OCHCH_2OH$ | **** | **** | 0.0366 | -19.2869 | **** | -0.6939 | -0.1438 | **** | **** | **** |
| 136 | $OCH_2CHOH$ | 0.1611 | **** | 0.0275 | -19.5555 | **** | 0.2807 | -0.1101 | **** | **** | **** |
| 137 | -O-OH | **** | **** | 0.1372 | -10.3102 | **** | **** | 0.6889 | **** | **** | **** |
| 138 | $CH_2SH$ | 1.5431 | 1.8996 | 0.4204 | -16.6831 | -2.2197 | 3.2668 | **** | **** | **** | **** |
| 139 | CHSH | 0.2299 | **** | 0.0476 | -16.7708 | **** | **** | **** | **** | **** | **** |
| 140 | CSH | **** | **** | -0.1342 | -14.8742 | **** | **** | **** | **** | **** | **** |
| 141 | aC-SH | **** | **** | 0.8198 | -16.6828 | **** | 3.3208 | **** | **** | **** | **** |
| 142 | -SH except as above | 0.7657 | **** | -0.0389 | -10.6945 | **** | 3.0292 | **** | **** | **** | **** |
| 143 | $CH_3S$ | 0.4408 | 0.4059 | 0.4553 | -16.4597 | 0.5000 | 2.6231 | **** | **** | **** | **** |
| 144 | $CH_2S$ | 0.6252 | 1.0444 | 0.3767 | -16.2943 | -0.4380 | **** | **** | **** | **** | **** |
| 145 | CHS | 0.6236 | -0.3983 | 0.5770 | -15.8430 | -1.3246 | -1.5055 | **** | **** | **** | **** |
| 146 | CS | 1.4408 | **** | 0.2262 | -16.8443 | **** | **** | **** | **** | **** | **** |
| 147 | aC-S- | 0.8459 | 0.3129 | 0.3336 | -15.7817 | -1.0938 | **** | **** | **** | **** | **** |
| 148 | SO | -1.3615 | -2.3995 | 0.1737 | -14.8690 | -1.4619 | 1.7578 | -0.5158 | **** | **** | **** |
| 149 | $SO_2$ | 2.5611 | **** | 0.3939 | -21.8506 | -0.5763 | 2.2691 | **** | **** | **** | **** |
| 150 | $SO_3$ (sulfite) | **** | **** | -0.2231 | -31.4604 | **** | **** | **** | **** | **** | **** |
| 151 | $SO_3$ (Sulfonate) | 0.4219 | **** | 0.3187 | -26.8991 | **** | **** | **** | **** | **** | **** |
| 152 | $SO_4$ (Sulfate) | **** | **** | 0.7967 | -32.6492 | **** | 2.8573 | **** | **** | **** | **** |
| 153 | aC-SO | -0.5944 | **** | 0.6778 | -21.4786 | -2.4005 | **** | **** | **** | **** | **** |
| 154 | aC-$SO_2$ | 0.2814 | -1.0660 | 0.0021 | -25.8954 | -0.3242 | **** | **** | **** | **** | **** |
| 155 | PH (phosphine) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 156 | P (Phospine) | **** | **** | 0.5950 | **** | **** | **** | **** | **** | **** | **** |
| 157 | $PO_3$ (Phospite) | **** | **** | 0.4576 | **** | -2.1138 | **** | **** | **** | **** | **** |
| 158 | $PHO_3$ (Phosponate) | 0.3208 | **** | 0.0665 | -25.9962 | **** | **** | **** | **** | **** | **** |
| 159 | $PO_3$ (Phosponate) | -0.0390 | 3.2035 | 1.4048 | -26.0106 | -1.5601 | **** | **** | **** | **** | **** |
| 160 | $PHO_4$ (Phospate) | 1.4767 | **** | 0.0500 | -30.9428 | -2.2271 | **** | **** | **** | **** | **** |
| 161 | $PO_4$ (Phospate) | 0.0857 | 1.6308 | 1.8110 | -30.7800 | -2.2559 | **** | **** | **** | **** | **** |
| 162 | aC-$PO_4$ | 1.9467 | 2.5522 | 1.1535 | -36.5009 | -1.3727 | -0.4513 | **** | **** | **** | **** |
| 163 | aC-P | **** | **** | 0.6937 | **** | **** | **** | **** | **** | **** | **** |
| 164 | $CO_3$ (Carbonate) | **** | **** | 0.4026 | -20.0556 | **** | **** | 1.5672 | **** | **** | **** |
| 165 | $C_2H_3O$ | 0.8514 | **** | 0.1473 | -14.0019 | **** | 1.0419 | 0.7183 | **** | **** | **** |
| 166 | $C_2H_2O$ | -0.4972 | **** | -0.0635 | -13.5542 | **** | **** | **** | **** | **** | **** |
| 167 | $C_2HO$ | **** | **** | -0.0608 | -13.9569 | **** | **** | **** | **** | **** | **** |
| 168 | $CH_2$ (cyclic) | 0.0897 | -0.0164 | 0.0305 | -4.9128 | 0.2381 | 0.2678 | 0.0522 | **** | **** | **** |
| 169 | CH (cyclic) | 0.3440 | 0.2230 | 0.1009 | -4.7455 | 0.0526 | -0.1033 | 0.0076 | **** | **** | **** |
| 170 | C (cyclic) | -0.0751 | 0.1635 | 0.2675 | -4.2661 | -0.4317 | -0.6719 | 0.0405 | **** | **** | **** |
| 171 | CH=CH (cyclic) | 0.4053 | 0.1881 | 0.0342 | -9.1455 | 0.8670 | 1.1926 | -0.1694 | **** | **** | **** |
| 172 | CH=C (cyclic) | 0.7101 | 0.6201 | 0.1961 | -9.0741 | -0.9429 | -0.3924 | -0.5956 | **** | **** | **** |
| 173 | C=C (cyclic) | 0.8299 | 0.9138 | 0.3052 | -9.1726 | -0.1101 | -0.7986 | **** | **** | **** | **** |
| 174 | $CH_2$=C (cyclic) | 1.3284 | 0.3457 | 0.6753 | -8.3601 | 1.2508 | 1.6196 | -0.4417 | **** | **** | **** |
| 175 | NH (cyclic) | -0.5359 | 0.0690 | 0.1039 | -5.0559 | -0.4237 | 3.5925 | **** | **** | **** | **** |
| 176 | N (cyclic) | -0.5352 | 0.4133 | 0.1509 | -4.1475 | -0.3082 | 0.2011 | -0.1283 | **** | **** | **** |
| 177 | CH=N (cyclic) | -0.1997 | 0.0016 | 0.0247 | -9.4054 | 0.0169 | **** | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 178 | C=N (cyclic) | -0.0297 | 1.6297 | 0.1815 | -9.3322 | -0.6622 | **** | **** | **** | **** | **** |
| 179 | O (cyclic) | -0.0178 | -0.3136 | 0.0485 | -5.2684 | -0.4086 | 1.0976 | 0.1839 | **** | **** | **** |
| 180 | CO (cyclic) | -0.0751 | -0.0348 | 0.0362 | -9.7030 | -0.2482 | 1.4906 | 0.1464 | **** | **** | **** |
| 181 | S (cyclic) | 0.5408 | -0.0221 | 0.0662 | -11.1395 | 0.2438 | -3.4289 | **** | **** | **** | **** |
| 182 | $SO_2$ (cyclic) | **** | -0.1274 | 0.2491 | -21.4443 | -0.5767 | **** | **** | **** | **** | **** |
| 183 | >NH | 0.0122 | -0.2127 | 0.2383 | -4.6764 | -0.6793 | **** | **** | **** | **** | **** |
| 184 | -O- | -1.0802 | 1.9031 | 0.3655 | -5.2880 | -0.2204 | -6.9696 | **** | **** | **** | **** |
| 185 | -S- | 0.3532 | 2.2864 | 0.0645 | -11.7688 | -0.1536 | **** | **** | **** | **** | **** |
| 186 | >CO | 1.1595 | 0.4557 | 0.0463 | -9.8009 | 0.4628 | 1.8939 | 0.1130 | **** | **** | **** |
| 187 | $PO_2$ | **** | **** | 0.1374 | **** | **** | **** | **** | **** | **** | **** |
| 188 | CH-N | **** | 1.7603 | -0.1090 | -8.9657 | **** | **** | **** | **** | **** | **** |
| 189 | SiHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 190 | SiO | **** | **** | -0.0007 | -16.0461 | -0.3101 | -3.1492 | **** | **** | **** | **** |
| 191 | $SiH_2$ | **** | **** | -0.2393 | **** | **** | **** | **** | **** | **** | **** |
| 192 | SiH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 193 | Si | **** | **** | 0.3458 | -11.2041 | -1.3606 | **** | **** | **** | **** | **** |
| 194 | $(CH_3)_3N$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 195 | N=N | 0.4933 | **** | 0.0539 | -9.3563 | **** | **** | **** | **** | **** | **** |
| 196 | $C_{cyclic}$=N- | 0.8793 | **** | 0.6162 | -9.3709 | -0.7675 | **** | **** | **** | **** | **** |
| 197 | $C_{cyclic}$=CH- | 2.3085 | 2.1495 | 0.6016 | -8.8475 | -0.6405 | 1.3880 | **** | **** | **** | **** |
| 198 | $C_{cyclic}$=NH | **** | **** | -0.0748 | -9.8549 | **** | **** | **** | **** | **** | **** |
| 199 | N=O | -0.3575 | -0.1965 | 0.3113 | -10.0720 | -0.3770 | **** | **** | **** | **** | -0.0775 |
| 200 | $C_{cyclic}$=C | 0.6756 | **** | 0.5676 | -8.5743 | -0.9050 | **** | -0.1664 | **** | **** | **** |
| 201 | P=O | 1.2123 | 2.2490 | 0.8100 | -16.0285 | -1.1477 | **** | **** | **** | **** | **** |
| 202 | N=N | **** | -1.2397 | 0.3611 | -10.7648 | -1.4104 | **** | **** | **** | **** | **** |
| 203 | C=NH | -0.4605 | **** | -0.1942 | -9.9767 | -0.8936 | **** | **** | **** | **** | **** |
| 204 | >C=S | 0.9273 | 0.3273 | 0.3458 | -15.7902 | 0.0321 | **** | **** | **** | **** | **** |
| 205 | aC-CON | -0.1061 | 2.2769 | 0.2972 | -17.6529 | -2.6869 | **** | **** | **** | **** | **** |
| 206 | aC=O | -0.3836 | **** | -0.0132 | -10.1596 | 0.1362 | **** | **** | **** | **** | **** |
| 207 | aN- | -0.6546 | 0.3255 | 0.1369 | -4.1070 | -1.0346 | **** | **** | **** | **** | **** |
| 208 | -Na | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 209 | -K | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 210 | HCONH | **** | **** | -0.1554 | -14.4278 | -0.7736 | **** | **** | **** | **** | **** |
| 211 | CHOCH | 1.2130 | 0.4709 | 0.3383 | -13.6062 | **** | **** | **** | **** | **** | **** |
| 212 | $C_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 213 | $SiH_3$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 214 | $SiH_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 215 | CH=C=CH | **** | **** | **** | -13.3744 | **** | **** | **** | **** | **** | **** |
| 216 | CH=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 217 | OP(=S)O | 0.8141 | 2.9977 | 1.4164 | -33.4285 | -0.2438 | 0.5000 | **** | **** | **** | **** |
| 218 | R | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 219 | $CF_{2cyclic}$ | **** | **** | 0.2788 | -17.8927 | **** | **** | **** | 1.0010 | **** | **** |
| 220 | $CF_{cyclic}$ | **** | **** | 0.3840 | -11.5240 | **** | **** | **** | **** | **** | **** |

[a] The symbols $LC50(FM)_{1i}$, $LC50(DM)_{1i}$, $LD50_{1i}$, $LogWs_{1i}$, $BCF_{1i,}$ $PEL_{1i,}$ $PCO_{1i}$, $GWP_{1i}$, $ODP_{1i}$,and $AP_{1i}$ represent the contributions ($C_i$) of the first-order groups for the corresponding properties.

**Table S2 (continued). MG Method Based Property Models Analysed Using Step-Wise Regression Method: First-Order Groups and their Contributions [a] for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| | Group | $EUA_{C\ 1i}$ | $EUA_{NC\ 1i}$ | $ERA_{C\ 1i}$ | $ERA_{NC\ 1i}$ | $EFW_{C\ 1i}$ | $EFW_{NC\ 1i}$ | $ESW_{C\ 1i}$ | $ESW_{NC\ 1i}$ | $ENS_{C\ 1i}$ | $ENS_{NC\ 1i}$ | $EAS_{C\ 1i}$ | $EAS_{NC\ 1i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $CH_3$ | 0.2812 | 0.0044 | 0.3053 | 0.1229 | 0.2626 | 0.1627 | 0.1362 | -0.1075 | 0.2433 | 0.0393 | 0.1771 | 0.0598 |
| 2 | $CH_2$ | -0.0243 | -0.0768 | -0.0173 | -0.0888 | -0.0503 | -0.1054 | -0.1035 | -0.1460 | 0.1514 | 0.0997 | 0.0888 | 0.0041 |
| 3 | CH | -0.2311 | 0.0074 | -0.0458 | -0.0536 | 0.1164 | -0.1383 | 0.1699 | -0.1374 | 0.2068 | 0.2690 | 0.0505 | 0.0314 |
| 4 | C | -0.2714 | -0.2538 | -0.8209 | -0.8273 | -0.5915 | -0.9984 | -0.6146 | -0.5315 | -0.6888 | -0.1677 | -0.7997 | -0.5965 |
| 5 | $CH_2$=CH | 0.4720 | -1.1792 | 0.7120 | -0.7414 | 0.3625 | -1.0817 | -0.1128 | -0.9796 | 0.4585 | -0.4659 | 0.3632 | -0.4702 |
| 6 | CH=CH | 0.0099 | -0.6462 | -0.0053 | -0.3547 | -0.5021 | -0.3094 | -0.6042 | -0.1727 | -0.2822 | -0.3880 | -0.3203 | -0.3861 |
| 7 | $CH_2$=C | -0.0002 | -0.3848 | 0.9076 | -0.0390 | 0.4974 | -0.3473 | -0.0990 | -0.1033 | 1.7056 | 0.4189 | 1.1123 | 0.1110 |
| 8 | CH=C | -0.4408 | 0.2765 | -0.1353 | -0.0375 | -0.3035 | -0.1969 | -0.5763 | 0.0312 | 0.0683 | 0.2096 | 0.2482 | -0.0689 |
| 9 | C=C | -0.3975 | -0.0285 | -0.9184 | -0.4819 | -0.8157 | -0.4467 | -1.4670 | -0.2164 | -0.4700 | -0.0352 | -0.5471 | -0.1193 |
| 10 | $CH_2$=C=CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | $CH_2$=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 12 | C=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | CH≡C | -1.4524 | -0.7521 | -0.6233 | -0.7808 | -0.6030 | -0.7958 | -0.0924 | 0.1487 | -0.5263 | -0.4364 | -0.6330 | -0.5975 |
| 14 | C≡C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 15 | aCH | 0.1490 | -0.0163 | 0.1566 | 0.0107 | 0.1167 | 0.0106 | 0.0640 | -0.0121 | 0.1571 | 0.0233 | 0.1393 | 0.0255 |
| 16 | aC fused with aromatic ring | -0.3459 | 0.0171 | -0.4712 | 0.0209 | -0.3477 | -0.0973 | -0.4930 | -0.2515 | -0.1678 | 0.2473 | -0.2880 | 0.1559 |
| 17 | aC fused with non-aromatic ring | -0.2140 | -0.0898 | -0.2943 | -0.1356 | -0.2763 | -0.1125 | -0.5172 | -0.3238 | -0.0410 | 0.2553 | -0.1375 | 0.1006 |
| 18 | aC except as above | -0.8266 | -0.0240 | -0.8915 | -0.3353 | -0.9367 | -0.3258 | -1.0364 | -0.9082 | -0.4495 | 0.2073 | -0.4613 | -0.0179 |
| 19 | aN in aromatic ring | 0.1644 | -0.3598 | -0.0029 | -0.3664 | 0.2228 | -0.1959 | 0.3106 | 0.1521 | 0.2173 | -0.4254 | 0.2630 | -0.3613 |
| 20 | aC-$CH_3$ | -0.2600 | -0.1517 | 0.1183 | 0.0806 | -0.1506 | 0.0707 | -0.2347 | -0.1264 | -0.0250 | 0.2765 | -0.1506 | 0.1845 |
| 21 | aC-$CH_2$ | -0.2461 | -0.0119 | -0.5533 | -0.0096 | -0.3687 | -0.0487 | -0.4396 | -0.0241 | -0.1477 | 0.6430 | -0.1678 | 0.3112 |
| 22 | aC-CH | -0.0544 | -0.2853 | -0.5470 | -0.5044 | -0.6417 | -0.7429 | -0.0676 | -0.3662 | -0.0490 | -0.3446 | 0.0021 | -0.4457 |
| 23 | aC-C | -0.6955 | -0.4066 | -1.1136 | -0.7185 | -1.2866 | -1.3071 | -1.6503 | -0.9373 | -0.8133 | -0.5220 | -0.8327 | -0.5737 |
| 24 | aC-CH=$CH_2$ | -0.0279 | -0.0406 | 0.4409 | 0.8714 | 1.6304 | 1.6651 | 0.1127 | 0.8952 | 0.8303 | 1.6498 | 1.1118 | 1.6201 |
| 25 | aC-CH=CH | 0.3962 | -0.6772 | -0.0339 | -1.2093 | -0.0013 | -0.9732 | -0.5925 | -2.3205 | 0.8344 | -0.5988 | 0.4860 | -0.6970 |
| 26 | aC-C=$CH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 27 | aC-C≡CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 28 | aC-C≡C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 29 | OH | 0.1973 | -0.1329 | 0.1224 | -0.0335 | 0.3576 | 0.1939 | 0.5909 | 0.4026 | 0.0373 | -0.3806 | 0.1133 | -0.2356 |
| 30 | aC-OH | 0.3579 | -0.2124 | 0.2272 | -0.3507 | 0.4162 | -0.1482 | 0.6553 | 0.4879 | 0.6292 | 0.3186 | 0.6065 | 0.1571 |
| 31 | COOH | 0.4136 | -0.4133 | 0.1754 | -0.7440 | 0.4426 | 0.0566 | 0.8145 | 1.3425 | 0.1276 | -0.4999 | 0.1350 | -0.6773 |
| 32 | aC-COOH | -0.4401 | 0.0701 | 0.3832 | -0.2474 | 0.4901 | 0.2991 | 1.1441 | 1.6362 | 0.6152 | 0.0521 | 0.3617 | -0.0445 |
| 33 | $CH_3CO$ | -0.5659 | 0.2394 | -1.0870 | 0.4385 | -0.7815 | 0.8784 | -0.8496 | 0.7182 | -0.5576 | 0.9002 | -0.7829 | 0.7620 |
| 34 | $CH_2CO$ | -1.2892 | **** | -1.4790 | **** | -0.6787 | **** | 0.1067 | **** | -2.2429 | **** | -1.9003 | **** |
| 35 | CHCO | **** | -0.4946 | **** | -0.6214 | **** | -0.2234 | **** | -0.0715 | **** | -0.2631 | **** | -0.5728 |
| 36 | CCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 37 | aC-CO | -0.6823 | -0.2301 | -1.7511 | -0.8199 | -0.8861 | -0.8717 | -0.2081 | -0.7125 | -1.3882 | 0.0514 | -1.3193 | -0.5369 |
| 38 | CHO | 0.9768 | -1.2611 | 0.6049 | -0.8794 | 0.7182 | -0.4964 | 0.7156 | -0.0583 | 0.5516 | -0.8968 | 0.5158 | -0.9705 |
| 39 | aC-CHO | 1.4448 | -0.1404 | 1.3160 | -0.1508 | 1.3270 | -0.1747 | 1.2618 | -0.2802 | 1.0550 | 0.4847 | 0.9880 | 0.0184 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | $CH_3COO$ | 0.5459 | 0.9003 | 0.4241 | 0.6052 | 0.7444 | 1.4409 | 0.3646 | 1.5768 | 0.2724 | 0.9531 | 0.2505 | 0.8098 |
| 41 | $CH_2COO$ | 0.6057 | 1.0128 | 0.0881 | 0.7938 | 0.5648 | 1.5247 | 0.9133 | 2.4827 | -0.3351 | 1.4263 | 0.1243 | 1.3393 |
| 42 | CHCOO | **** | -0.0308 | **** | -0.4444 | **** | 1.0059 | **** | 0.4328 | **** | 1.3017 | **** | 0.2341 |
| 43 | CCOO | -0.8771 | **** | -1.1997 | **** | -1.4991 | **** | -1.2924 | **** | -1.7785 | **** | -1.8290 | **** |
| 44 | HCOO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 45 | aC-COO | -0.3185 | 0.2823 | -0.4614 | -0.0342 | 0.0421 | 0.3270 | -0.0685 | 0.7594 | 0.2976 | 0.5649 | 0.0071 | 0.2475 |
| 46 | aC-OOCH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 47 | aC-OOC | 0.6800 | 0.8397 | 1.1171 | 0.2293 | 4.3179 | 0.6441 | 7.4590 | 0.3575 | 3.2164 | 1.1469 | 1.2900 | 0.6910 |
| 48 | COO except as above | -0.1330 | 0.0421 | -0.5584 | -0.1768 | -0.0621 | 0.3225 | 0.5514 | 0.7978 | -0.2484 | 0.6865 | -0.1042 | 0.3132 |
| 49 | $CH_3O$ | -0.0737 | 0.3926 | 0.0890 | 0.3819 | 0.3565 | 0.2597 | -0.0584 | 0.3986 | -0.0537 | -0.1449 | 0.1894 | 0.0078 |
| 50 | $CH_2O$ | 0.5146 | 0.3708 | 0.5731 | 0.1905 | 0.2700 | 0.2248 | 0.1110 | 0.2236 | 0.2602 | -0.2207 | 0.1187 | -0.1002 |
| 51 | CH-O | **** | 0.2217 | **** | 0.2546 | **** | -0.1904 | **** | 0.8482 | **** | 0.2456 | **** | -0.1258 |
| 52 | C-O | 0.7438 | **** | 0.3865 | **** | 0.9454 | **** | 2.4004 | **** | 0.9286 | **** | 0.1394 | **** |
| 53 | aC-O | -0.1124 | -0.2269 | -0.1606 | -0.3367 | -0.1500 | -0.4353 | 0.1490 | -0.1858 | -0.0825 | -0.1939 | -0.1134 | -0.1872 |
| 54 | $CH_2NH_2$ | 2.6417 | **** | 2.2328 | **** | 2.3637 | **** | 2.6710 | **** | 1.9022 | **** | 2.1752 | **** |
| 55 | $CHNH_2$ | -0.4403 | **** | -0.4944 | **** | -0.6307 | **** | 0.2293 | **** | -0.9306 | **** | -0.6118 | **** |
| 56 | $CNH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 57 | $CH_3NH$ | 0.5989 | 1.1599 | 1.2116 | 0.8850 | 0.1976 | 1.7821 | 0.3662 | 1.8835 | 0.2348 | 2.1442 | 0.2376 | 1.6882 |
| 58 | $CH_2NH$ | -0.2935 | 2.6870 | 0.2179 | 2.9085 | -0.4139 | 1.6683 | -0.3472 | 0.4786 | -0.3451 | 2.2910 | -0.2829 | 2.6186 |
| 59 | CHNH | **** | -1.0359 | **** | -0.8527 | **** | -1.5413 | **** | -0.8592 | **** | -1.9285 | **** | -1.9808 |
| 60 | $CH_3N$ | -0.9628 | 0.5031 | -0.7580 | -0.0732 | -0.8942 | 0.0312 | -0.7265 | 0.5274 | -0.7908 | 0.1335 | -0.7504 | 0.2359 |
| 61 | $CH_2N$ | -0.0865 | -2.9995 | -0.1413 | -2.5338 | -0.3720 | -3.8875 | -0.1841 | -3.2261 | -0.2994 | -3.0576 | -0.1682 | -2.5588 |
| 62 | aC-$NH_2$ | 0.3571 | -0.2153 | 0.2748 | -0.3095 | 0.1398 | -0.2644 | 0.4196 | -0.0304 | 0.2464 | -0.1121 | 0.2767 | -0.1759 |
| 63 | aC-NH | -0.0048 | 0.2582 | -0.2628 | -0.1291 | -0.2814 | -0.2055 | -0.1881 | -0.0031 | -0.2280 | 0.1322 | -0.2492 | 0.0722 |
| 64 | aC-N | -0.4604 | 1.0680 | -0.4588 | 0.7264 | -0.9315 | 0.3995 | -0.9075 | 0.1869 | -0.8443 | 0.5837 | -0.9327 | 0.2939 |
| 65 | $NH_2$ except as above | -0.1142 | 0.0985 | -0.4335 | 0.0042 | -0.2549 | 0.2402 | 0.0728 | 1.3282 | -0.3473 | 0.4343 | -0.3593 | 0.2852 |
| 66 | CH=N | 0.0044 | -0.4784 | 0.3889 | -0.7022 | 0.3618 | -0.9722 | 0.3871 | -0.5138 | 0.7647 | -0.5477 | 0.8108 | -0.7022 |
| 67 | C=N | 0.3823 | 0.1238 | -0.0164 | 0.0187 | 0.2669 | -0.6474 | 0.1658 | -0.0183 | 1.0080 | 0.5949 | 0.9758 | 0.3875 |
| 68 | $CH_2CN$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 69 | CHCN | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 70 | CCN | -1.8216 | **** | -1.9527 | **** | -1.7907 | **** | -1.5534 | **** | -2.2276 | **** | -2.1245 | **** |
| 71 | aC-CN | 1.2567 | -0.1886 | 0.9475 | -0.2222 | 1.2346 | -0.0152 | 1.5580 | 0.2173 | 1.3322 | -0.1325 | 1.3219 | 0.0023 |
| 72 | CN except as above | 0.0264 | -0.1436 | -0.7595 | -0.1718 | -0.1286 | 0.3670 | 0.0422 | 0.0000 | -0.1534 | 0.0266 | -0.2672 | -0.0738 |
| 73 | $CH_2NCO$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 74 | CHNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 75 | CNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 76 | aC-NCO | 1.1089 | -1.9959 | 0.5349 | -2.1334 | 0.4265 | -2.0027 | -0.2182 | -2.4080 | 0.9661 | -1.5508 | 0.8283 | -1.7168 |
| 77 | $CH2NO_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 78 | $CHNO_2$ | -0.9658 | -0.9958 | -1.7870 | -1.4643 | -1.0063 | -0.6800 | -2.0912 | -0.9445 | -1.5018 | -0.5350 | -1.1962 | -0.6291 |
| 79 | $CNO_2$ | 0.7609 | **** | 1.1691 | **** | 0.4194 | **** | -1.2488 | **** | 0.7154 | **** | 0.7829 | **** |
| 80 | aC-$NO_2$ | 0.0987 | -0.7413 | -0.5705 | -0.9110 | -0.1609 | -0.6097 | -0.1511 | -0.4803 | -0.0347 | -0.3346 | 0.0709 | -0.4266 |
| 81 | $NO_2$ except as above | -0.6643 | 0.2770 | -1.0546 | 0.2802 | -0.6147 | 0.5173 | -0.5777 | 0.5233 | -0.7462 | 1.3837 | -0.7208 | 1.1761 |
| 82 | ONO | 0.6316 | **** | -0.0875 | **** | 0.2952 | **** | -0.7739 | **** | -0.2319 | **** | 0.2707 | **** |
| 83 | $ONO_2$ | 0.4802 | **** | 0.0600 | **** | 0.2000 | **** | 0.3815 | **** | 0.0576 | **** | 0.1529 | **** |
| 84 | $HCON(CH_2)_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 85 | $HCONHCH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | $CONH_2$ | 0.0134 | -0.4139 | 0.1691 | -0.0586 | 0.8177 | -0.1241 | 1.4121 | 0.8899 | 0.7314 | -0.5006 | 0.8945 | 0.0155 |
| 87 | $CONHCH_3$ | -0.0313 | -0.8362 | -0.7594 | -1.0911 | -0.4951 | -0.8312 | -0.1024 | 0.2246 | -0.8420 | -0.8833 | -0.5860 | -1.0145 |
| 88 | $CONHCH_2$ | 1.9518 | 1.1442 | 1.4832 | 1.4330 | 1.2426 | 1.1152 | 1.5296 | 1.6594 | 2.2234 | 2.1503 | 2.1266 | 1.9013 |
| 89 | $CON(CH_3)_2$ | -0.6537 | -0.0950 | -1.0803 | -0.3867 | -1.1343 | -0.1322 | -0.1937 | 0.6511 | -1.3163 | -0.2674 | -1.2917 | -0.3364 |
| 90 | $CONCH_3CH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 91 | $CON(CH_2)_2$ | -0.8273 | -0.3936 | -0.9693 | -0.7407 | -1.3969 | -0.8424 | -1.1112 | -0.1034 | -1.1740 | -0.5455 | -0.9545 | -0.8783 |
| 92 | CONHCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 93 | CONCO | **** | -0.9068 | **** | -1.4827 | **** | -0.8258 | **** | 0.3186 | **** | -1.0391 | **** | -0.8294 |
| 94 | aC-$CONH_2$ | 0.8871 | **** | -0.0865 | **** | 0.8239 | **** | 0.5860 | **** | 0.2734 | **** | 0.4742 | **** |
| 95 | aC-NH(CO)H | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 96 | aC-N(CO)H | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 97 | aC-CONH | -0.7583 | 1.2697 | -1.4627 | 0.6663 | -0.6019 | 1.0400 | -0.2033 | 2.0214 | -1.3456 | 1.0294 | -0.8983 | 0.6978 |
| 98 | aC-NHCO | -0.0004 | -0.0637 | -0.2983 | -0.5320 | -0.1121 | 0.0350 | 0.3804 | 0.1428 | -0.4114 | 0.2488 | -0.2086 | 0.1182 |
| 99 | aC-(N)CO | -1.0967 | 0.3045 | -1.4759 | -0.3026 | -1.3678 | -0.4157 | -0.5218 | 0.6672 | -0.6634 | -0.2291 | -0.4365 | -0.2363 |
| 100 | NHCONH | 1.4303 | **** | 0.9832 | **** | 1.1946 | **** | 1.9429 | **** | 0.4362 | **** | 0.5252 | **** |
| 101 | $NH_2CONH$ | 0.3821 | **** | -0.3543 | **** | 0.0467 | **** | 0.7847 | **** | -0.0753 | **** | 0.0593 | **** |
| 102 | $NH_2CON$ | -0.4982 | **** | -1.1906 | **** | -0.7382 | **** | -0.0828 | **** | -0.7912 | **** | -0.7436 | **** |
| 103 | NHCON | 0.3024 | 0.0100 | -0.5509 | -0.7257 | 0.2361 | 0.2184 | 0.6711 | 1.3846 | 0.3913 | -0.6951 | 0.2303 | -0.4350 |
| 104 | NCON | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 105 | aC-$NHCONH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 106 | aC-NHCONH | 1.4478 | 1.0207 | 1.2121 | 0.5931 | 0.9761 | 0.4899 | 1.8983 | 0.5083 | 1.4380 | -0.0717 | 1.3208 | 0.0327 |
| 107 | NHCO except as above | 0.1199 | -0.0146 | 0.0454 | 0.0032 | 0.2878 | 0.2213 | 0.8196 | 0.3913 | 0.2920 | 0.0295 | 0.3624 | 0.0287 |
| 108 | $CH_2Cl$ | -0.3830 | -0.7636 | -0.3819 | -0.6860 | -0.3472 | -0.3273 | -0.4563 | -0.8939 | -0.3392 | -0.6522 | -0.4056 | -0.5439 |
| 109 | CHCl | 0.6637 | 0.1485 | -0.2285 | -0.2019 | 0.5142 | -0.6331 | -0.9817 | -0.9698 | 0.1789 | -0.2755 | 0.2965 | 0.1149 |
| 110 | CCl | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 111 | $CHCl_2$ | 0.4531 | -0.7610 | -0.1357 | -1.1547 | 0.3226 | -0.5988 | -0.6482 | -1.1267 | 0.0549 | -0.6492 | 0.0719 | -0.6546 |
| 112 | $CCl_2$ | **** | 0.0162 | **** | -0.1154 | **** | -0.4194 | **** | -1.0026 | **** | 0.1964 | **** | 0.0306 |
| 113 | $CCl_3$ | -0.0116 | -1.2569 | -0.6886 | -1.3607 | -0.3901 | -1.2703 | -1.5385 | -1.5906 | -0.6039 | -0.9174 | -0.5410 | -0.9872 |
| 114 | $CH_2F$ | 1.8052 | 1.9109 | 1.2625 | 1.2313 | 1.6943 | 2.6100 | 1.0269 | 1.5339 | 1.2637 | 2.2373 | 1.3172 | 1.8204 |
| 115 | CHF | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 116 | CF | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 117 | $CHF_2$ | **** | -0.0027 | **** | 0.1968 | **** | 0.2337 | **** | -0.6164 | **** | 0.8043 | **** | 0.3066 |
| 118 | $CF_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 119 | $CF_3$ | 0.7993 | -0.4068 | 0.0981 | -0.4311 | 0.5850 | -0.6634 | -0.8071 | -1.1163 | 0.1960 | -0.8828 | 0.3345 | -0.4122 |
| 120 | $CCl_2F$ | **** | -1.4244 | **** | -2.2499 | **** | -1.2401 | **** | -2.6623 | **** | -1.6824 | **** | -1.6484 |
| 121 | HCClF | **** | 1.3816 | **** | 0.4334 | **** | 1.5902 | **** | 0.3557 | **** | 1.1147 | **** | 1.1246 |
| 122 | $CClF_2$ | **** | 0.9918 | **** | 1.0841 | **** | 1.2244 | **** | 1.0152 | **** | 1.0564 | **** | 1.0765 |
| 123 | aC-Cl | -0.1496 | -0.2926 | -0.2917 | -0.4226 | -0.3034 | -0.3287 | -0.7378 | -0.7018 | -0.3028 | -0.2349 | -0.2799 | -0.2913 |
| 124 | aC-F | -0.5147 | -0.6887 | -1.2858 | -0.5469 | -0.8683 | -0.4136 | -1.0689 | -0.4146 | -0.8543 | -0.2960 | -1.0622 | -0.3001 |
| 125 | aC-I | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 126 | aC-Br | 0.1178 | -0.2363 | -0.0015 | -0.2642 | 0.1150 | -0.1578 | -0.0861 | -0.2954 | 0.1426 | -0.1230 | 0.0672 | -0.1986 |
| 127 | -I except as above | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 128 | -Br except as above | -0.0869 | -0.1776 | -0.0596 | -0.2436 | -0.2059 | 0.2137 | -0.3641 | -0.1295 | -0.3791 | -0.1064 | -0.2836 | -0.0557 |
| 129 | -F except as above | 0.3896 | -0.1807 | 0.3099 | 0.0680 | 0.4146 | 0.1057 | 0.3737 | 0.0762 | 0.2175 | -0.0558 | 0.2341 | -0.0119 |
| 130 | -Cl except as above | 0.0452 | -0.2286 | -0.0167 | -0.2860 | -0.0168 | -0.0381 | -0.1609 | -0.3444 | -0.0951 | -0.2142 | -0.0896 | -0.2182 |
| 131 | CHNOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 132 | CNOH | 0.0478 | **** | -0.7777 | **** | -0.3279 | **** | 0.1303 | **** | -0.1868 | **** | 0.0328 | **** |
| 133 | aC-CHNOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 134 | $OCH_2CH_2OH$ | 2.1707 | -0.1215 | 1.7390 | -0.2071 | 1.8167 | -0.0192 | 2.7585 | 1.5986 | 1.3109 | -0.1664 | 1.3937 | -0.2047 |
| 135 | $OCHCH_2OH$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 136 | $OCH_2CHOH$ | **** | 1.0611 | **** | 1.0028 | **** | 0.9576 | **** | 2.7896 | **** | 1.1904 | **** | 1.0355 |
| 137 | -O-OH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 138 | $CH_2SH$ | -0.7931 | **** | -1.8639 | **** | -1.3399 | **** | -0.5572 | **** | -0.2854 | **** | -0.2384 | **** |
| 139 | CHSH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 140 | CSH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 141 | aC-SH | -1.6674 | **** | -3.2165 | **** | -2.0555 | **** | -0.7549 | **** | -2.3760 | **** | -2.0847 | **** |
| 142 | -SH except as above | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 143 | $CH_3S$ | 1.7560 | -1.1500 | -0.1111 | -0.4163 | 0.0458 | -0.3930 | 0.4226 | -1.2235 | 0.1624 | -1.4553 | -0.0226 | -1.2049 |
| 144 | $CH_2S$ | -0.3952 | -0.2160 | -0.9617 | -0.2021 | -1.1017 | -0.3669 | -1.3556 | -0.6114 | -0.9489 | -0.2602 | -0.9053 | -0.4426 |
| 145 | CHS | **** | -0.1118 | **** | -0.3187 | **** | -1.2181 | **** | -1.6430 | **** | -1.7585 | **** | -1.3933 |
| 146 | CS | 1.6458 | -0.0620 | 1.4751 | 0.5075 | 1.2477 | -1.0085 | 0.4709 | -0.6864 | 1.0890 | -0.8233 | 1.1165 | -0.8747 |
| 147 | aC-S- | -1.1541 | -0.2388 | -1.6190 | -0.6546 | -1.3791 | -0.6194 | -1.5459 | -0.9009 | -1.1808 | -0.2158 | -1.2017 | -0.4052 |
| 148 | SO | 1.4991 | 1.2728 | 0.5711 | 1.4900 | 0.2190 | 1.3598 | 0.4583 | 0.6590 | 0.2160 | 1.7927 | 0.4250 | 1.1434 |
| 149 | SO2 | **** | -0.7598 | **** | -1.1765 | **** | -0.3042 | **** | -0.4479 | **** | -0.0085 | **** | -0.0109 |
| 150 | $SO_3$ (sulfite) | 0.6697 | 0.6069 | -0.1826 | 0.1647 | -0.5130 | 0.3638 | -0.8241 | -0.7865 | -0.2758 | 0.2651 | 0.0493 | 0.1334 |
| 151 | $SO_3$ (Sulfonate) | -0.3168 | **** | -1.3506 | **** | -0.5966 | **** | 0.4606 | **** | -0.8646 | **** | -0.6894 | **** |
| 152 | $SO_4$ (Sulfate) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 153 | aC-SO | **** | -1.5062 | **** | -1.6110 | **** | -1.3267 | **** | -0.0555 | **** | -1.4480 | **** | -1.2697 |
| 154 | aC-$SO_2$ | 0.2941 | -0.6688 | -0.4563 | -1.0517 | 0.0824 | -0.7098 | 0.5214 | -0.4493 | -0.2997 | 0.0352 | -0.1496 | -0.2397 |
| 155 | PH (phosphine) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 156 | P (Phospine) | **** | -1.2820 | **** | -1.5965 | **** | -2.3001 | **** | -1.6462 | **** | -0.6856 | **** | -0.0428 |
| 157 | $PO_3$ (Phospite) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 158 | $PHO_3$ (Phosponate) | 0.7416 | **** | -0.0104 | **** | 0.3076 | **** | 1.1965 | **** | 0.0906 | **** | 0.2658 | **** |
| 159 | $PO_3$ (Phosponate) | 1.8953 | -0.4482 | 1.3651 | -1.2315 | 1.2858 | -0.4895 | 2.0993 | 1.0570 | 1.4417 | -0.6521 | 1.5191 | -0.1747 |
| 160 | $PHO_4$ (Phospate) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 161 | $PO_4$ (Phospate) | 1.0798 | -2.2565 | -0.2379 | -2.1906 | 0.7802 | -2.1982 | 1.4306 | -0.5733 | 0.9302 | -2.1455 | 0.6254 | -1.8969 |
| 162 | aC-$PO_4$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 163 | aC-P | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 164 | $CO_3$ (Carbonate) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 165 | $C_2H_3O$ | 0.4798 | -1.4409 | -0.0067 | -1.6703 | 0.4151 | -1.0459 | -0.2880 | -1.2669 | -0.3115 | -1.0362 | 0.0989 | -1.1662 |
| 166 | $C_2H_2O$ | **** | -0.2548 | **** | -0.2048 | **** | -0.4602 | **** | -0.3366 | **** | -0.9231 | **** | -0.9071 |
| 167 | $C_2HO$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 168 | $CH_2$ (cyclic) | 0.0517 | 0.0477 | 0.0258 | 0.0949 | -0.0231 | 0.0852 | -0.0035 | 0.0641 | 0.1168 | 0.1538 | 0.0651 | 0.1094 |
| 169 | CH (cyclic) | -0.2099 | -0.1818 | -0.3139 | -0.2209 | -0.2499 | -0.3282 | -0.4216 | -0.3743 | -0.1332 | -0.0170 | -0.1400 | -0.0728 |
| 170 | C (cyclic) | -0.2783 | -0.1269 | -0.3415 | -0.1110 | -0.3659 | -0.3299 | -0.4277 | -0.1314 | -0.1758 | -0.0743 | -0.1693 | -0.0609 |
| 171 | CH=CH (cyclic) | 0.2143 | -0.3422 | 0.1756 | -0.2311 | -0.2314 | -0.2424 | -0.1734 | -0.5072 | 0.2430 | 0.2965 | 0.1448 | 0.1844 |
| 172 | CH=C (cyclic) | 0.3631 | 0.1375 | 0.2563 | 0.1158 | -0.0462 | 0.0647 | -0.0189 | -0.0488 | 0.4705 | 0.3370 | 0.2823 | 0.1311 |
| 173 | C=C (cyclic) | -0.1056 | -0.7511 | -0.2382 | -0.9559 | -0.4197 | -1.3426 | -0.3613 | -1.0790 | 0.1772 | -0.9770 | 0.1060 | -0.8705 |
| 174 | $CH_2$=C (cyclic) | -0.1864 | **** | -0.5550 | **** | -1.2561 | **** | -1.4313 | **** | 0.0896 | **** | 0.3598 | **** |
| 175 | NH (cyclic) | 0.0173 | -0.2067 | -0.0522 | -0.3755 | 0.0650 | -0.2712 | 0.2082 | 0.2561 | 0.0309 | -0.5038 | 0.0956 | -0.3555 |
| 176 | N (cyclic) | 0.0067 | -0.3155 | -0.1687 | -0.4413 | -0.0415 | -0.5931 | 0.2860 | -0.1578 | -0.3214 | -1.3094 | -0.1979 | -1.0750 |
| 177 | CH=N (cyclic) | 0.8193 | 0.1688 | 0.6106 | 0.1964 | 0.6874 | 0.3340 | 0.9412 | 0.2484 | 0.3549 | 0.6275 | 0.3681 | 0.7590 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 178 | C=N (cyclic) | -0.2137 | -0.3953 | -0.3968 | -0.4652 | -0.1178 | -0.3734 | 0.1774 | -0.1057 | -0.1374 | 0.0737 | 0.0007 | 0.0397 |
| 179 | O (cyclic) | -0.5237 | -0.1128 | -0.3063 | -0.1283 | -0.2134 | -0.1151 | -0.1357 | 0.0910 | -0.7957 | -0.5423 | -0.6411 | -0.4295 |
| 180 | CO (cyclic) | 0.0827 | 0.2883 | -0.0578 | 0.1261 | 0.1300 | 0.4942 | 0.4265 | 0.6506 | 0.0014 | 0.6848 | 0.0099 | 0.5919 |
| 181 | S (cyclic) | -0.2378 | 0.0549 | -0.2629 | 0.3156 | -0.1562 | 0.0932 | -0.2200 | 0.3792 | 0.0032 | 0.0763 | 0.0118 | -0.0247 |
| 182 | $SO_2$ (cyclic) | 0.1035 | -0.1536 | -0.7377 | -0.4337 | 0.9810 | -0.2873 | 1.0942 | 0.1567 | 3.4223 | -0.5213 | 3.3165 | -0.5782 |
| 183 | >NH | -0.4571 | -0.0645 | -0.3987 | -0.3324 | -0.2798 | -0.3977 | -0.5169 | -0.7340 | -0.2534 | -0.7876 | -0.2042 | -0.6020 |
| 184 | -O- | 1.7348 | 0.1749 | 0.3176 | -0.5076 | 0.1044 | 0.5573 | -0.6604 | 0.6133 | 0.2070 | 0.5392 | -0.6565 | 0.6999 |
| 185 | -S- | 1.5902 | 0.3574 | 1.3400 | 0.3992 | 0.9095 | 0.2659 | 0.5240 | 0.1034 | 1.7518 | -0.2023 | 1.6974 | -0.0040 |
| 186 | >CO | 0.0857 | -0.5621 | **** | -0.8385 | -0.6724 | -0.4349 | -0.5629 | -0.2143 | -0.2026 | -0.3192 | **** | -0.4260 |
| 187 | $PO_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 188 | CH-N | **** | 0.2157 | -0.7919 | 0.2827 | **** | -0.7992 | **** | 0.1344 | **** | 0.2756 | -0.6340 | 0.0122 |
| 189 | SiHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 190 | SiO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 191 | $SiH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 192 | SiH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 193 | Si | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 194 | $(CH_3)_3N$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 195 | N=N | **** | 0.0102 | **** | -0.1583 | **** | 0.0432 | **** | 0.1089 | **** | -0.4251 | **** | -0.1029 |
| 196 | $C_{cyclic}$=N- | -0.5479 | -0.8507 | -0.7442 | -0.7851 | -0.1413 | -0.3556 | -0.6733 | -0.2041 | 0.3919 | 0.9717 | 0.1476 | 0.5806 |
| 197 | $C_{cyclic}$=CH- | -0.3670 | **** | -0.6552 | **** | -0.5085 | **** | -0.6474 | **** | -0.9219 | **** | -0.7721 | **** |
| 198 | $C_{cyclic}$=NH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 199 | N=O | -0.9060 | **** | -0.8677 | **** | -0.9591 | **** | -0.6917 | **** | -1.1569 | **** | -1.0287 | **** |
| 200 | $C_{cyclic}$=C | -0.7306 | **** | -1.0157 | **** | -1.0434 | **** | -1.0337 | **** | -1.0385 | **** | -0.9418 | **** |
| 201 | P=O | -0.6489 | -1.7203 | -0.0055 | -1.9083 | -0.0211 | -1.6554 | 0.6175 | -0.5475 | -0.1279 | -0.8717 | -0.0973 | -1.3289 |
| 202 | N=N | 0.3441 | **** | -0.6377 | **** | -0.1406 | **** | -0.6545 | **** | -0.6320 | **** | -0.6844 | **** |
| 203 | C=NH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 204 | >C=S | 0.4296 | -1.0799 | 0.1655 | -0.5831 | 0.0996 | -0.8105 | 0.5208 | -0.5155 | -0.0525 | 0.0162 | -0.0481 | -0.3127 |
| 205 | aC-CON | -0.2883 | -1.6096 | -0.9420 | -2.0087 | -0.7021 | -2.6680 | 0.4358 | -3.0736 | -0.4751 | -2.7795 | -0.4727 | -1.9244 |
| 206 | aC=O | 2.0541 | **** | 2.0441 | **** | 2.1876 | **** | 2.6538 | **** | 2.4073 | **** | 2.2166 | **** |
| 207 | aN- | -0.7235 | -0.8058 | -0.5085 | -1.0327 | -0.5252 | -0.7093 | -0.0804 | 0.2873 | -0.0185 | -0.0801 | 0.3597 | -0.0003 |
| 208 | -Na | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 209 | -K | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 210 | HCONH | 0.6315 | **** | 0.1483 | **** | 0.5060 | **** | 1.0373 | **** | 0.1419 | **** | 0.1964 | **** |
| 211 | CHOCH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 212 | $C_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 213 | $SiH_3$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 214 | $SiH_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 215 | CH=C=CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 216 | CH=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 217 | OP(=S)O | 0.9193 | -1.0765 | 0.8755 | -1.0305 | -0.0055 | -1.6725 | 1.4290 | -1.1229 | 0.1428 | -0.8306 | -0.2776 | -1.1215 |
| 218 | R | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 219 | $CF_{2cyclic}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 220 | $CF_{cyclic}$ | -1.1221 | **** | -1.5038 | **** | -1.3502 | **** | -2.3291 | **** | -0.9893 | **** | -1.0702 | **** |

[a] The symbols $EUA_{C\ 1i}$, $EUA_{NC\ 1i}$, $ERA_{C\ 1i}$, $ERA_{NC\ 1i}$, $EFW_{C\ 1i}$, $EFW_{NC\ 1i}$, $ESW_{C\ 1i}$, $ESW_{NC\ 1i}$, $ENS_{C\ 1i}$, $ENS_{NC\ 1i}$, $EAS_{C\ 1i}$, and $EAS_{NC\ 1i}$ represent the contributions ($C_i$) of the first-order groups for the corresponding properties.

**Table S3. MG Method Based Property Models Analysed Using Step-Wise Regression Method: Second-Order Groups and their Contributions [a] for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), and PCO**

| | Group | $LC50(FM)_{2j}$ | $LC50(DM)_{2j}$ | $LD50_{2j}$ | $LogWs_{2j}$ | $BCF_{2j}$ | $PEL_{2j}$ | $PCO_{2j}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | $(CH_3)_2CH$ | -0.0764 | -0.1057 | -0.0308 | -0.0961 | 0.1168 | -0.0213 | 0.0154 |
| 2 | $(CH_3)_3C$ | 0.0412 | 0.0599 | 0.0840 | -0.3055 | -0.0811 | 0.0227 | 0.0272 |
| 3 | $CH(CH_3)CH(CH_3)$ | -0.1113 | **** | 0.0468 | -0.1821 | 1.0401 | 0.0043 | 0.0580 |
| 4 | $CH(CH_3)C(CH_3)_2$ | **** | **** | -0.1376 | 0.4430 | 0.8297 | 0.0346 | -0.0299 |
| 5 | $C(CH_3)_2C(CH_3)_2$ | **** | **** | 0.0640 | -0.5905 | **** | -0.0239 | 0.0148 |
| 6 | $CH_n$=$CH_m$-$CH_p$=$CH_k$ (*k,m,n,p* in 0..2) | -0.3839 | 0.4580 | 0.0901 | -0.7096 | -0.5140 | 0.2521 | 0.3386 |
| 7 | $CH_3$-$CH_m$=$CH_n$ (*m,n* in 0..2) | 0.2163 | -1.5389 | -0.0529 | -0.0474 | 0.2127 | 0.3417 | -0.0467 |
| 8 | $CH_2$-$CH_m$=$CH_n$ (*m,n* in 0..2) | -0.2321 | 0.1522 | -0.0234 | -0.0205 | 0.0272 | 0.2745 | -0.0203 |
| 9 | $CH_p$-$CH_m$=$CH_n$ (*m,n* in 0..2; *p* in 0..1) | 0.2029 | -0.3690 | 0.1804 | 0.1311 | -0.1108 | **** | -0.0588 |
| 10 | CHCHO or CCHO | 0.1432 | **** | 0.0030 | -0.3170 | **** | **** | -0.1355 |
| 11 | $CH_3COCH_2$ | -0.0733 | 0.2120 | -0.0806 | 0.1427 | **** | -0.0934 | -0.0677 |
| 12 | $CH_3COCH$ or $CH_3COC$ | 0.4132 | **** | 0.2043 | 0.5001 | 0.0430 | -0.1821 | 0.0025 |
| 13 | CHCOOH or CCOOH | -0.6512 | **** | 0.1233 | 0.3928 | -0.2597 | 0.5903 | 0.0367 |
| 14 | $CH_3COOCH$ or $CH_3COOC$ | -0.0341 | **** | 0.0951 | 0.1888 | **** | -0.1284 | -0.0162 |
| 15 | CO-O-CO | **** | **** | -0.0807 | -0.6322 | 0.0000 | 1.0520 | **** |
| 16 | CHOH | 0.0554 | 1.0759 | 0.0219 | 0.4393 | -0.2086 | 0.0954 | 0.0132 |
| 17 | COH | 0.0625 | **** | 0.0891 | 0.7215 | -0.1750 | -0.1548 | 0.3195 |
| 18 | $CH_3COCH_nOH$ (*n* in 0..2) | **** | **** | -0.0982 | 0.8644 | **** | **** | -0.1783 |
| 19 | NCCHOH or NCCOH | **** | **** | 0.1495 | 0.6017 | **** | 0.0000 | **** |
| 20 | OH-$CH_n$-COO (*n* in 0..2) | 0.8775 | **** | -0.0409 | 0.2457 | -0.4364 | **** | -0.2612 |
| 21 | $CH_m(OH)CH_n(OH)$ (*m,n* in 0..2) | -1.2012 | -1.1768 | 0.0066 | -0.1728 | -0.5661 | 0.2618 | -0.0693 |
| 22 | $CH_m(OH)CH_n(NH_p)$ (*m,n,p* in 0..2) | -0.1316 | 0.6018 | -0.0205 | 0.2429 | 0.0000 | 0.3565 | -0.0042 |
| 23 | $CH_m(NH_2)CH_n(NH_2)$ (*m,n* in 0..2) | 0.0265 | 0.0777 | -0.4353 | -0.7650 | **** | -0.4550 | **** |
| 24 | $CH_m(NH)CH_n(NH_2)$ (*m,n* in 1..2) | **** | **** | -0.1929 | 0.3688 | 0.0000 | -0.3635 | **** |
| 25 | $H_2NCOCH_nCH_mCONH_2$ (*m,n* in 1..2) | **** | **** | **** | -0.6827 | **** | **** | **** |
| 26 | $CH_m(NH_n)$-COOH (*m,n* in 0..2) | 0.4575 | 0.6214 | -0.1690 | -0.1789 | **** | **** | **** |
| 27 | HOOC-$CH_n$-COOH (*n* in 1..2) | **** | **** | -0.1155 | 0.5780 | **** | **** | **** |
| 28 | HOOC-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | 0.0445 | 0.2286 | 1.8872 | **** | -0.3516 |
| 29 | HO-$CH_n$-COOH (*n* in 1..2) | **** | **** | -0.0777 | -0.3265 | 0.2712 | **** | 0.1355 |
| 30 | $NH_2$-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | -0.6598 | **** | **** | **** |
| 31 | $CH_3$-O-$CH_n$-COOH (n in 1..2) | **** | **** | **** | 0.7930 | **** | **** | **** |
| 32 | HS-CH-COOH | 0.0000 | **** | 0.6687 | 0.2806 | **** | -1.2326 | **** |
| 33 | HS-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | 0.0364 | 0.8204 | **** | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 34 | NC-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | 0.5699 | 0.1470 | **** | -0.4976 | **** |
| 35 | OH-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | 0.2069 | 0.9963 | **** | **** | **** |
| 36 | HS-$CH_n$-$CH_m$-SH (*n, m* in 1..2) | **** | **** | -1.5941 | **** | **** | **** | **** |
| 37 | COO-$CH_n$-$CH_m$-OOC (*n, m* in 1..2) | **** | **** | 0.1041 | -0.3202 | **** | **** | -0.1422 |
| 38 | OOC-$CH_m$-$CH_m$-COO (*n, m* in 1..2) | -0.1100 | -0.1899 | -0.0098 | 0.3351 | 0.0000 | 0.0000 | 0.3766 |
| 39 | NC-$CH_n$-COO (*n* in 1..2) | **** | **** | -0.2553 | 0.0320 | **** | **** | **** |
| 40 | $COCH_n$COO (*n* in 1..2) | **** | **** | 0.0455 | 0.2919 | **** | **** | **** |
| 41 | $CH_m$-O-$CH_n$=$CH_p$ (*m,n,p* in 0..3) | **** | **** | -0.3996 | -0.0117 | 1.1102 | **** | **** |
| 42 | $CH_m$=$CH_n$-F (*m,n* in 0..2) | **** | **** | 0.0100 | -0.3430 | **** | -0.0351 | **** |
| 43 | $CH_m$=$CH_n$-Br (*m,n* in 0..2) | **** | -0.0478 | -0.3156 | 0.3983 | **** | 0.8884 | **** |
| 44 | $CH_m$=$CH_n$-I (*m,n* in 0..2) | **** | **** | **** | 0.0121 | **** | **** | **** |
| 45 | $CH_m$=$CH_n$-Cl (*m,n* in 0..2) | 0.0205 | 0.0144 | 0.1063 | 0.1214 | 0.0502 | -0.1321 | -0.0613 |
| 46 | $CH_m$=$CH_n$-CN (*m,n* in 0..2) | -0.2490 | 0.4082 | 0.1708 | 1.0178 | **** | -0.1118 | 0.0000 |
| 47 | $CH_n$=$CH_m$-COO-$CH_p$ (*m,n,p* in 0..3) | 0.0570 | 0.4720 | -0.0032 | 0.2879 | **** | -0.2021 | 0.0143 |
| 48 | $CH_m$=$CH_n$-CHO (*m,n* in 0..2) | 1.3458 | 2.0090 | 0.2139 | 0.4115 | **** | 0.1212 | 0.2420 |
| 49 | $CH_m$=$CH_n$-COOH (*m,n* in 0..2) | 1.0013 | **** | -0.0370 | 0.0592 | **** | -0.5149 | -0.2860 |
| 50 | aC-$CH_n$-X (*n* in 1..2) X: Halogen | -0.0077 | **** | -0.0723 | -0.0254 | -0.2510 | 0.2746 | **** |
| 51 | aC-$CH_n$-$NH_m$ (*n* in 1..2; *m* in 0..2)) | -0.1092 | **** | -0.1196 | 0.5674 | 0.0255 | **** | **** |
| 52 | aC-$CH_n$-O- (*n* in 1..2) | 0.3513 | -0.1209 | 0.1138 | 0.5302 | -0.0011 | **** | **** |
| 53 | aC-$CH_n$-OH (*n* in 1..2) | -0.1727 | **** | -0.0304 | 0.1802 | -0.0459 | **** | -0.0430 |
| 54 | aC-$CH_n$-CN (*n* in 1..2) | 1.4486 | -0.1985 | 0.0300 | 0.5431 | -0.7076 | **** | **** |
| 55 | aC-$CH_n$-CHO (*n* in 1..2) | **** | **** | -0.2336 | -0.2078 | **** | **** | **** |
| 56 | aC-$CH_n$-SH (*n* in 1..2) | **** | **** | 0.2557 | **** | **** | **** | **** |
| 57 | aC-$CH_n$-COOH (*n* in 1..2) | **** | **** | 0.3508 | 0.1899 | -0.5435 | **** | **** |
| 58 | aC-$CH_n$-CO- (*n* in 1..2) | 0.0000 | **** | 0.0819 | 0.3811 | -0.3243 | **** | **** |
| 59 | aC-$CH_n$-S- (*n* in 1..2) | **** | **** | -0.4356 | 0.1455 | -0.0417 | **** | **** |
| 60 | aC-$CH_n$-OOC-H (*n* in 1..2) | **** | **** | 0.1375 | **** | **** | **** | **** |
| 61 | aC-$CH_m$-$NO_2$ (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** |
| 62 | aC-$CH_n$-$CONH_2$ (*n* in 1..2) | **** | **** | 0.2234 | 0.4008 | **** | **** | **** |
| 63 | aC-$CH_n$-OOC (*n* in 1..2) | -0.2284 | -0.6714 | -0.1266 | -0.7736 | -0.0125 | 0.5009 | **** |
| 64 | aC-$CH_n$-COO (*n* in 1..2) | -0.8038 | 0.5838 | 0.2030 | -0.1084 | 0.4011 | **** | **** |
| 65 | aC-$SO_2$-OH | -0.0528 | **** | -0.0221 | 0.6120 | 0.0638 | **** | **** |
| 66 | aC-C$H(CH_3)_2$ | -0.1479 | -0.0398 | -0.2498 | -0.3480 | 0.0743 | 0.3027 | 0.0372 |
| 67 | aC-C$(CH_3)_3$ | 0.4288 | **** | 0.0038 | -0.0489 | -0.0393 | 0.4519 | 0.0160 |
| 68 | aC-$CF_3$ | -0.0138 | -0.0943 | -0.0622 | 0.0992 | 0.0751 | **** | 0.0111 |
| 69 | $(CH_n$=C$)_{cyc}$-CHO (*n* in 0..2) | **** | **** | -0.1006 | -0.3013 | **** | -0.5824 | **** |
| 70 | $(CH_n$=C$)_{cyc}$-COO-$CH_m$ (*n,m* in 0..3) | **** | **** | -0.0304 | 0.2300 | -0.0373 | **** | **** |
| 71 | $(CH_n$=C$)_{cyc}$-CO- (*n* in 0..2) | -0.5436 | **** | 0.0327 | -0.2010 | **** | **** | **** |
| 72 | $(CH_n$=C$)_{cyc}$-$CH_3$ (*n* in 0..2) | -0.1629 | 0.2697 | -0.0287 | 0.1151 | -0.1004 | -0.2472 | -0.1472 |
| 73 | $(CH_n$=C$)_{cyc}$-$CH_2$ (*n* in 0..2) | 0.0686 | -2.1549 | 0.0605 | 0.4049 | **** | -0.6411 | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 74 | $(CH_n{=}C)_{cyc}$-CN ($n$ in 0..2) | -0.6072 | **** | -0.6863 | -0.2613 | **** | **** | **** |
| 75 | $(CH_n{=}C)_{cyc}$-Cl ($n$ in 0..2) | 0.1507 | 0.0423 | -0.0377 | 0.0344 | 0.0199 | 0.1901 | **** |
| 76 | $CH_{cyc}$-$CH_3$ | -0.0418 | 0.0679 | 0.0533 | 0.0404 | -0.1411 | -0.2157 | -0.0279 |
| 77 | $CH_{cyc}$-$CH_2$ | -0.5577 | -1.2746 | 0.0237 | 0.1272 | 0.2565 | -0.2110 | -0.0028 |
| 78 | $CH_{cyc}$-CH | 0.7411 | -2.6422 | -0.0129 | 0.0975 | -0.2238 | **** | 0.0378 |
| 79 | $CH_{cyc}$-C | -0.4529 | **** | -0.2425 | -0.2294 | 0.4370 | **** | -0.1253 |
| 80 | $CH_{cyc}$-CH=$CH_n$ ($n$ in 1..2) | 0.3363 | -2.3849 | 0.1087 | -0.3632 | -0.1528 | 1.2756 | **** |
| 81 | $CH_{cyc}$-C=$CH_n$ ($n$ in 1..2) | -0.7578 | 2.1395 | -0.1690 | 0.0967 | -1.1805 | **** | 0.6709 |
| 82 | $CH_{cyc}$-Cl | -0.1520 | -0.0695 | 0.1234 | 0.0072 | -0.0429 | -1.1409 | **** |
| 83 | $CH_{cyc}$-F | **** | **** | **** | -0.0294 | **** | **** | **** |
| 84 | $CH_{cyc}$-OH | -0.1258 | 0.6398 | 0.0353 | -0.1178 | -0.3225 | -0.3254 | -0.0604 |
| 85 | $CH_{cyc}$-$NH_2$ | **** | **** | -0.1432 | 0.1418 | **** | 0.1139 | **** |
| 86 | $CH_{cyc}$-NH-$CH_n$ ($n$ in 0..3) | 0.2855 | 0.3464 | -0.1079 | 0.4598 | -0.0967 | **** | **** |
| 87 | $CH_{cyc}$-N-$CH_n$ ($n$ in 0..3) | 0.7025 | **** | -0.1946 | 0.3802 | 0.4684 | **** | **** |
| 88 | $CH_{cyc}$-SH | **** | **** | 1.4912 | **** | **** | 0.4203 | **** |
| 89 | $CH_{cyc}$-CN | **** | **** | -0.3849 | 0.7628 | **** | **** | **** |
| 90 | $CH_{cyc}$-COOH | -0.2236 | **** | 0.0332 | -0.1894 | 0.0867 | **** | **** |
| 91 | $CH_{cyc}$-CO | **** | **** | 0.1850 | 0.1888 | **** | **** | **** |
| 92 | $CH_{cyc}$-$NO_2$ | **** | **** | 0.5918 | 0.7474 | **** | **** | **** |
| 93 | $CH_{cyc}$-S- | -1.1224 | **** | 0.0123 | 0.6629 | **** | **** | **** |
| 94 | $CH_{cyc}$-CHO | -0.0488 | **** | -0.1554 | -0.1686 | **** | **** | -0.0594 |
| 95 | $CH_{cyc}$-O- | **** | **** | -0.0867 | 0.0702 | **** | 1.5492 | **** |
| 96 | $CH_{cyc}$-OOCH | **** | **** | 0.1650 | **** | **** | **** | **** |
| 97 | $CH_{cyc}$-COO | -0.4375 | 3.9708 | -0.1197 | -0.1930 | 0.5332 | **** | **** |
| 98 | $CH_{cyc}$-OOC | -0.9174 | 0.8222 | -0.0947 | 0.2024 | **** | **** | **** |
| 99 | $C_{cyc}$-$CH_3$ | 0.0618 | -0.2693 | -0.0245 | -0.1054 | 0.0856 | 0.4295 | 0.0161 |
| 100 | $C_{cyc}$-$CH_2$ | 0.5211 | 0.8205 | 0.0520 | 0.2880 | -0.1757 | 0.5001 | -0.0554 |
| 101 | $C_{cyc}$-OH | -0.4611 | 0.9501 | 0.2298 | -0.1743 | -0.9768 | -0.8443 | **** |
| 102 | >$N_{cyc}$-$CH_3$ | 0.4824 | 0.2009 | 0.0145 | 0.1060 | -0.3611 | **** | -0.1030 |
| 103 | >$N_{cyc}$-$CH_2$ | -0.1675 | 0.0000 | 0.0030 | 0.2220 | 0.1704 | **** | **** |
| 104 | AROMRINGs$^1$s$^2$ | 0.1461 | 0.1005 | 0.0332 | -0.0987 | 0.0801 | -0.0995 | -0.0705 |
| 105 | AROMRINGs$^1$s$^3$ | -0.0819 | 0.0190 | 0.0470 | -0.2251 | 0.0114 | -0.2389 | -0.1364 |
| 106 | AROMRINGs$^1$s$^4$ | 0.0545 | 0.1187 | -0.0160 | -0.1804 | -0.0626 | 0.1977 | -0.0255 |
| 107 | AROMRINGs$^1$s$^2$s$^3$ | -0.1163 | 0.1842 | 0.1096 | -0.0196 | 0.1842 | 0.2225 | -0.1012 |
| 108 | AROMRINGs$^1$s$^2$s$^4$ | 0.0807 | -0.0351 | -0.0055 | -0.0269 | -0.0430 | -0.1142 | 0.0171 |
| 109 | AROMRINGs$^1$s$^3$s$^5$ | -0.1749 | -0.1048 | 0.0383 | -0.3208 | 0.4586 | 0.2071 | -0.1189 |
| 110 | AROMRINGs$^1$s$^2$s$^3$s$^4$ | 0.2641 | -0.2547 | -0.1754 | 0.0873 | -0.0073 | **** | 0.0249 |
| 111 | AROMRINGs$^1$s$^2$s$^3$s$^5$ | -0.1944 | 0.0897 | -0.0836 | 0.0750 | 0.0998 | -1.1173 | 0.1087 |
| 112 | AROMRINGs$^1$s$^2$s$^4$s$^5$ | 0.1374 | -0.1737 | -0.0633 | -0.1630 | 0.1481 | **** | 0.0211 |
| 113 | PYRIDINEs$^2$ | -0.5596 | **** | 0.3831 | 0.7714 | -0.0366 | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 114 | PYRIDINEs$^{3}$ | -0.1207 | **** | -0.0853 | 0.7486 | -0.8944 | **** | **** |
| 115 | PYRIDINEs$^{4}$ | -0.3869 | **** | 0.2497 | 0.3145 | -0.2412 | **** | **** |
| 116 | PYRIDINEs$^{2}$s$^{3}$ | -0.8534 | **** | 0.4131 | 0.3683 | **** | **** | **** |
| 117 | PYRIDINEs$^{2}$s$^{4}$ | **** | **** | 0.1771 | 0.5459 | **** | **** | **** |
| 118 | PYRIDINEs$^{2}$s$^{5}$ | -0.2390 | **** | 0.0415 | 0.1791 | 1.4201 | **** | **** |
| 119 | PYRIDINEs$^{2}$s$^{6}$ | 0.0206 | **** | 0.3342 | -0.1606 | **** | **** | **** |
| 120 | PYRIDINEs$^{3}$s$^{4}$ | **** | **** | 0.2221 | 0.1017 | **** | **** | **** |
| 121 | PYRIDINEs$^{3}$s$^{5}$ | **** | **** | -0.1887 | -0.1816 | **** | **** | **** |
| 122 | PYRIDINEs$^{2}$s$^{3}$s$^{6}$ | **** | **** | 0.1607 | -0.2204 | **** | **** | **** |
| 123 | (CH$^{n}$=CH$^{m}$)$^{cyc}$-COOH | **** | **** | -0.8910 | -0.4421 | **** | **** | **** |
| 124 | AROMRINGs$^{1}$s$^{2}$s$^{3}$s$^{4}$s$^{5}$ | -0.1316 | -0.4815 | 0.0886 | 0.3891 | -0.0740 | **** | 0.2418 |
| 125 | aC-NHCOCH$_2$N | **** | **** | 0.6234 | 1.2707 | **** | **** | **** |
| 126 | (N=C)$_{cyc}$-CH$_3$ | -0.6390 | **** | 0.0287 | 0.4988 | -0.4374 | **** | **** |
| 127 | aC-CONH(CH$_2$)$_2$N | **** | **** | 0.3162 | 0.6654 | **** | **** | **** |
| 128 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | **** | **** | -0.0116 | -0.1908 | -0.3573 | **** | **** |
| 129 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | **** | **** | 0.3337 | 0.3478 | **** | **** | **** |
| 130 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | **** | **** | -0.1481 | -0.1134 | -0.1719 | **** | **** |

[a] The symbols LC50(FM)$_{2j}$, LC50(DM)$_{2j}$LD50$_{2j}$, LogWs$_{2j}$, BCF$_{2j,}$ PEL$_{2j,}$ PCO$_{2j}$ represent the contributions ($D_j$) of the second-order groups for the corresponding properties. Note that there are no second-order contributions for GWP, ODP and AP since these properties belong to chemicals whose structure is defined by first-order groups alone.

**Table S3 (continued). MG Method Based Property Models Analysed Using Step-Wise Regression Method: Second-Order Groups and their Contributions [a] for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| | Group | $EUA_{C\ 2j}$ | $EUA_{NC\ 2j}$ | $ERA_{C\ 2j}$ | $ERA_{NC2j}$ | $EFW_{C\ 2j}$ | $EFW_{NC\ 2j}$ | $ESW_{C\ 2j}$ | $ESW_{NC\ 2j}$ | $ENS_{C\ 2j}$ | $ENS_{NC\ 2j}$ | $EAS_{C\ 2j}$ | $EAS_{NC\ 2j}$ | $EUA_{C\ 2j}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $(CH_3)_2CH$ | 0.0493 | 0.1243 | -0.0350 | 0.0690 | -0.0267 | 0.1142 | 0.1385 | 0.1442 | -0.0370 | 0.1454 | -0.0149 | 0.1514 | 0.0493 |
| 2 | $(CH_3)_3C$ | -0.2156 | -0.0108 | 0.0723 | 0.0970 | -0.1280 | -0.0997 | -0.6370 | -0.0211 | -0.2256 | 0.0162 | 0.0317 | -0.1962 | -0.2156 |
| 3 | $CH(CH_3)CH(CH_3)$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 4 | $CH(CH_3)C(CH_3)_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 5 | $C(CH_3)_2C(CH_3)_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 6 | $CH_n$=$CH_m$-$CH_p$=$CH_k$ (*k,m,n,p* in 0..2) | 0.2736 | -0.0427 | -0.3420 | - | -0.5695 | -0.2080 | -0.2673 | -0.1188 | -0.1908 | -0.3873 | -0.2523 | -0.2213 | 0.2736 |
| 7 | $CH_3$-$CH_m$=$CH_n$ (*m,n* in 0..2) | -0.1712 | -0.0231 | 0.2339 | 0.0173 | 0.2614 | 0.0113 | 0.2153 | -0.0860 | 0.2118 | 0.1161 | 0.4139 | -0.0296 | -0.1712 |
| 8 | $CH_2$-$CH_m$=$CH_n$ (*m,n* in 0..2) | -0.3032 | 0.1616 | -0.0889 | 0.7139 | -0.3033 | 0.3887 | -0.5355 | 0.3290 | -0.2445 | 0.3041 | -0.3880 | 0.2059 | -0.3032 |
| 9 | $CH_p$-$CH_m$=$CH_n$ (*m,n* in 0..2; *p* in 0..1) | 0.2543 | -0.5451 | -0.2249 | - | 0.1906 | -0.6964 | 0.0219 | -0.3310 | -0.0087 | -0.6388 | -0.2203 | -0.5026 | 0.2543 |
| 10 | CHCHO or CCHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | $CH_3COCH_2$ | 0.3730 | -0.1831 | 0.3565 | 0.0548 | 0.4955 | -0.0975 | 0.8558 | 0.0463 | 0.0577 | -0.0729 | 0.2147 | -0.1488 | 0.3730 |
| 12 | $CH_3COCH$ or $CH_3COC$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | CHCOOH or CCOOH | 0.0349 | -0.2373 | -0.1692 | - | -0.0496 | -0.1837 | 0.0025 | -0.4877 | -0.0280 | -0.1855 | -0.2519 | -0.1631 | 0.0349 |
| 14 | $CH_3COOCH$ or $CH_3COOC$ | -0.0031 | **** | 0.1004 | **** | -0.6116 | **** | -0.2886 | **** | -0.1639 | **** | -0.0906 | **** | -0.0031 |
| 15 | CO-O-CO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 16 | CHOH | -0.0202 | -0.2409 | 0.0717 | - | -0.2283 | -0.3003 | -0.0856 | -0.2054 | 0.0897 | -0.1083 | 0.0650 | -0.1746 | -0.0202 |
| 17 | COH | 0.2691 | -0.3445 | -0.3730 | - | -0.1990 | -0.3745 | -0.1655 | -1.0201 | 0.2900 | -0.4718 | 0.1695 | -0.1091 | 0.2691 |
| 18 | $CH_3COCH_nOH$ (*n* in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 19 | NCCHOH or NCCOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 20 | OH-$CH_n$-COO (*n* in 0..2) | -0.1302 | -0.0724 | -0.9562 | - | 0.0508 | -0.3016 | -1.0696 | -0.2605 | -0.6475 | -0.0184 | -1.2065 | 0.1851 | -0.1302 |
| 21 | $CH_m(OH)CH_n(OH)$ (*m,n* in 0..2) | -0.6337 | **** | 0.4786 | 0.0674 | -2.0125 | **** | -0.0382 | **** | -0.3093 | **** | -0.2551 | **** | -0.6337 |
| 22 | $CH_m(OH)CH_n(NH_p)$ (*m,n,p* in 0..2) | -0.6256 | **** | -0.9371 | **** | -0.7020 | **** | -0.8702 | **** | -0.9641 | **** | -1.0751 | **** | -0.6256 |
| 23 | $CH_m(NH_2)CH_n(NH_2)$ (*m,n* in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 24 | $CH_m(NH)CH_n(NH_2)$ (*m,n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 25 | $H_2NCOCH_nCH_mCONH_2$ (*m,n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 26 | $CH_m(NH_n)$-COOH (*m,n* in 0..2) | 0.2262 | 0.0000 | 0.3603 | 0.0000 | 0.3488 | 0.0000 | 0.0218 | **** | 0.4185 | 0.0000 | 0.5089 | 0.0000 | 0.2262 |
| 27 | HOOC-$CH_n$-COOH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 28 | HOOC-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 29 | HO-$CH_n$-COOH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 30 | $NH_2$-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 31 | $CH_3$-O-$CH_n$-COOH (n in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 32 | HS-CH-COOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | HS-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 34 | NC-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 35 | OH-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 36 | HS-$CH_n$-$CH_m$-SH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 37 | COO-$CH_n$-$CH_m$-OOC (*n, m* in 1..2) | 0.4469 | **** | 0.6427 | **** | 0.6875 | **** | 0.0983 | **** | 0.4387 | **** | 0.4819 | **** | 0.4469 |
| 38 | OOC-$CH_m$-$CH_m$-COO (*n, m* in 1..2) | -0.6981 | 0.4660 | -0.7761 | 0.6166 | -0.7811 | 0.8178 | -0.0708 | 0.5411 | 0.3322 | 1.2502 | -0.3033 | 1.3833 | -0.6981 |
| 39 | NC-$CH_n$-COO (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 40 | $COCH_nCOO$ (*n* in 1..2) | -1.8650 | **** | -1.7823 | **** | -2.4776 | **** | -4.2789 | **** | -0.2887 | **** | -1.0733 | **** | -1.8650 |
| 41 | $CH_m$-O-$CH_n$=$CH_p$ (*m,n,p* in 0..3) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 42 | $CH_m$=$CH_n$-F (*m,n* in 0..2) | -0.0771 | **** | -0.0802 | **** | -0.0851 | **** | -0.0019 | **** | -0.0470 | **** | -0.0282 | **** | -0.0771 |
| 43 | $CH_m$=$CH_n$-Br (*m,n* in 0..2) | **** | -0.1566 | **** | - | **** | -0.3653 | **** | -0.3419 | **** | -0.2270 | **** | -0.0672 | **** |
| 44 | $CH_m$=$CH_n$-I (*m,n* in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 45 | $CH_m$=$CH_n$-Cl (*m,n* in 0..2) | 0.1340 | -0.0433 | 0.1338 | 0.0066 | 0.3215 | -0.0546 | 0.1440 | -0.1301 | 0.2013 | -0.0126 | 0.2409 | -0.0067 | 0.1340 |
| 46 | $CH_m$=$CH_n$-CN (*m,n* in 0..2) | -0.3279 | -0.7947 | -0.1976 | - | -0.4389 | -0.7243 | -0.9403 | -0.8521 | -0.6087 | -1.0236 | -0.5363 | -0.9052 | -0.3279 |
| 47 | $CH_n$=$CH_m$-COO-$CH_p$ (*m,n,p* in 0..3) | 0.0069 | 0.0317 | -0.2782 | - | -0.2485 | -0.2269 | 0.1015 | 0.0910 | -0.1050 | -0.0816 | 0.0177 | 0.0795 | 0.0069 |
| 48 | $CH_m$=$CH_n$-CHO (*m,n* in 0..2) | -0.6932 | -1.5263 | -0.0810 | - | 0.0021 | -0.6885 | 0.3819 | -2.2257 | -0.3543 | -1.2512 | -0.5143 | -1.1984 | -0.6932 |
| 49 | $CH_m$=$CH_n$-COOH (*m,n* in 0..2) | -0.8306 | **** | -0.8969 | **** | -1.0459 | **** | -1.2008 | **** | -0.9549 | **** | -0.8379 | **** | -0.8306 |
| 50 | aC-$CH_n$-X (*n* in 1..2) X: Halogen | -0.5178 | **** | -0.6314 | **** | -0.3923 | **** | -0.7383 | **** | -0.5454 | **** | -0.5294 | **** | -0.5178 |
| 51 | aC-$CH_n$-$NH_m$ (*n* in 1..2; *m* in 0..2)) | 0.4471 | **** | 0.2560 | **** | 0.1222 | **** | -0.0642 | **** | 0.3025 | **** | 0.5023 | **** | 0.4471 |
| 52 | aC-$CH_n$-O- (*n* in 1..2) | 0.6104 | 0.1555 | 0.4508 | 0.1158 | 0.3795 | -0.2851 | 0.5760 | -0.4067 | 0.3126 | 0.2177 | 0.4127 | 0.1494 | 0.6104 |
| 53 | aC-$CH_n$-OH (*n* in 1..2) | 0.1408 | -0.0105 | 0.4880 | - | 0.3236 | 0.0274 | 0.5686 | 0.1099 | 0.3018 | -0.0632 | 0.3113 | -0.1477 | 0.1408 |
| 54 | aC-$CH_n$-CN (*n* in 1..2) | **** | 0.0051 | **** | 0.1590 | **** | 0.0594 | **** | -0.0081 | **** | 0.2819 | **** | 0.0526 | **** |
| 55 | aC-$CH_n$-CHO (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 56 | aC-$CH_n$-SH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 57 | aC-$CH_n$-COOH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 58 | aC-$CH_n$-CO- (*n* in 1..2) | **** | -0.0357 | **** | 0.0104 | **** | 0.2390 | **** | 0.1591 | **** | 0.0396 | **** | 0.0493 | **** |
| 59 | aC-$CH_n$-S- (*n* in 1..2) | **** | -0.2919 | **** | - | **** | -0.1592 | **** | -0.2046 | **** | 0.3179 | **** | 0.3396 | **** |
| 60 | aC-$CH_n$-OOC-H (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 61 | aC-$CH_m$-$NO_2$ (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 62 | aC-$CH_n$-$CONH_2$ (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 63 | aC-$CH_n$-OOC (*n* in 1..2) | 0.4686 | 0.0303 | 0.4189 | 0.0462 | 0.3224 | 0.0302 | 0.5022 | 0.1718 | 0.3665 | -0.2482 | 0.3405 | -0.0679 | 0.4686 |
| 64 | aC-$CH_n$-COO (*n* in 1..2) | 0.3346 | -0.0228 | 0.5036 | 0.1422 | 0.0694 | -0.1243 | 0.3202 | -0.2149 | 0.4888 | -0.3434 | 0.4867 | -0.4429 | 0.3346 |
| 65 | aC-$SO_2$-OH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 66 | aC-$CH(CH_3)_2$ | -1.3901 | -0.0910 | -1.8195 | - | -1.4011 | 0.0001 | -0.4869 | 0.2649 | -1.1552 | -0.5553 | -0.9422 | -0.0808 | -1.3901 |
| 67 | aC-$C(CH_3)_3$ | 0.2991 | -0.0127 | 0.2530 | 0.2438 | 0.3429 | 0.0812 | 0.2798 | 1.0165 | 0.3542 | 0.5184 | 0.4483 | -0.2791 | 0.2991 |
| 68 | aC-$CF_3$ | 0.1531 | 0.0424 | 0.3470 | 0.1010 | 0.2803 | 0.1118 | -0.0413 | 0.1208 | 0.0848 | 0.0624 | 0.0273 | -0.0074 | 0.1531 |
| 69 | $(CH_n=C)_{cyc}$-CHO (*n* in 0..2) | 1.0675 | 0.3007 | 0.2864 | 0.4844 | 0.3154 | -0.2639 | 0.2947 | 0.6183 | 0.6517 | -0.1476 | 0.6216 | -0.0205 | 1.0675 |
| 70 | $(CH_n=C)_{cyc}$-COO-$CH_m$ (*n,m* in 0..3) | **** | 0.3055 | **** | 0.2118 | **** | 0.1074 | **** | 0.1839 | **** | -0.4792 | **** | -0.4464 | **** |
| 71 | $(CH_n=C)_{cyc}$-CO- (*n* in 0..2) | **** | 0.5239 | **** | 0.3025 | **** | 0.2088 | **** | 0.0438 | **** | 0.5817 | **** | 0.4229 | **** |
| 72 | $(CH_n=C)_{cyc}$-$CH_3$ (*n* in 0..2) | -0.0655 | 0.0845 | 0.3278 | 0.2269 | 0.4787 | 0.2541 | -0.0549 | 0.6388 | 0.2507 | 0.2678 | 0.3649 | 0.2391 | -0.0655 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 73 | $(CH_n{=}C)_{cyc}$-$CH_2$ (*n* in 0..2) | 0.4720 | -0.3818 | 0.0573 | - | 0.1077 | -0.5961 | 0.4593 | -0.6096 | -0.4267 | -0.2819 | -0.2841 | -0.3409 | 0.4720 |
| 74 | $(CH_n{=}C)_{cyc}$-CN (*n* in 0..2) | **** | 0.1111 | **** | 0.0100 | **** | -0.1014 | **** | -0.2254 | **** | 0.0394 | **** | 0.2240 | **** |
| 75 | $(CH_n{=}C)_{cyc}$-Cl (*n* in 0..2) | -0.0174 | 0.0085 | -0.0040 | 0.0654 | -0.2488 | 0.0744 | -0.3374 | -0.0197 | -0.1947 | 0.0863 | -0.1816 | 0.0594 | -0.0174 |
| 76 | $CH_{cyc}$-$CH_3$ | 0.1617 | -0.0418 | -0.1039 | - | -0.0965 | -0.0070 | -0.0103 | -0.0280 | -0.1309 | 0.1500 | -0.0806 | -0.1868 | 0.1617 |
| 77 | $CH_{cyc}$-$CH_2$ | 0.0635 | 0.2944 | 0.1360 | 0.3422 | 0.0175 | 0.3337 | -0.0708 | 0.5201 | 0.1780 | 0.3558 | 0.1664 | 0.2557 | 0.0635 |
| 78 | $CH_{cyc}$-CH | **** | -0.2875 | **** | 0.9491 | **** | -0.1082 | **** | 1.4752 | **** | 0.5631 | **** | 0.0253 | **** |
| 79 | $CH_{cyc}$-C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 80 | $CH_{cyc}$-$CH{=}CH_n$ (*n* in 1..2) | 1.3088 | 0.2967 | 1.7672 | 0.9563 | 1.1725 | 0.4304 | 1.1241 | 0.5865 | 0.9590 | 0.6805 | 0.5345 | 0.6288 | 1.3088 |
| 81 | $CH_{cyc}$-$C{=}CH_n$ (*n* in 1..2) | 0.2406 | 0.5031 | **** | 0.3233 | -1.3843 | 0.7125 | -0.4182 | 1.0504 | **** | 0.4853 | -2.2284 | 0.4768 | 0.2406 |
| 82 | $CH_{cyc}$-Cl | 0.0257 | -0.0812 | -0.0001 | - | 0.0151 | -0.1684 | 0.0731 | 0.0380 | -0.0066 | -0.1851 | 0.0001 | -0.1387 | 0.0257 |
| 83 | $CH_{cyc}$-F | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 84 | $CH_{cyc}$-OH | 0.0797 | **** | 0.2227 | **** | 0.5691 | **** | 0.3085 | **** | 0.6193 | **** | 0.4963 | **** | 0.0797 |
| 85 | $CH_{cyc}$-$NH_2$ | **** | -0.1407 | **** | 0.5338 | **** | -0.1351 | **** | -0.3309 | **** | -0.6419 | **** | -0.4300 | **** |
| 86 | $CH_{cyc}$-NH-$CH_n$ (*n* in 0..3) | **** | -0.1128 | **** | - | **** | -0.2112 | **** | -0.1903 | **** | -0.0236 | **** | -0.2425 | **** |
| 87 | $CH_{cyc}$-N-$CH_n$ (*n* in 0..3) | -0.0580 | **** | -0.2114 | **** | -1.0213 | **** | -0.8909 | **** | -0.3087 | **** | -1.0499 | **** | -0.0580 |
| 88 | $CH_{cyc}$-SH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 89 | $CH_{cyc}$-CN | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 90 | $CH_{cyc}$-COOH | 0.2481 | **** | 0.3285 | **** | 0.6706 | **** | 1.2973 | **** | 0.7755 | **** | 0.6862 | **** | 0.2481 |
| 91 | $CH_{cyc}$-CO | 0.8042 | **** | 0.8784 | **** | 1.6918 | **** | 1.6952 | **** | 1.6515 | **** | 1.9332 | **** | 0.8042 |
| 92 | $CH_{cyc}$-$NO_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 93 | $CH_{cyc}$-S- | **** | 0.4048 | **** | 0.1410 | **** | 0.4632 | **** | 0.3856 | **** | 0.3079 | **** | 0.3541 | **** |
| 94 | $CH_{cyc}$-CHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 95 | $CH_{cyc}$-O- | 0.6481 | 0.0368 | 0.6856 | - | -0.3891 | -0.0508 | 0.0841 | -0.5279 | -0.0898 | -0.2893 | -0.0996 | 0.0577 | 0.6481 |
| 96 | $CH_{cyc}$-OOCH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 97 | $CH_{cyc}$-COO | -1.4180 | 0.2074 | -0.9837 | - | 0.3256 | 0.5251 | 0.6104 | 0.1570 | 0.5905 | 0.2617 | 0.7028 | 0.0426 | -1.4180 |
| 98 | $CH_{cyc}$-OOC | -0.5002 | **** | -0.6494 | **** | 0.0684 | **** | -0.5904 | **** | -0.6940 | **** | -0.6726 | **** | -0.5002 |
| 99 | $C_{cyc}$-$CH_3$ | 0.4925 | 0.0098 | 0.3342 | 0.0934 | 0.2086 | -0.0724 | 0.3966 | 0.0352 | 0.1006 | -0.1272 | 0.1656 | -0.0368 | 0.4925 |
| 100 | $C_{cyc}$-$CH_2$ | -0.1094 | -0.4090 | -0.1517 | - | -0.1147 | -0.4371 | -0.0568 | -0.6990 | -0.2259 | -0.0504 | -0.2193 | -0.0427 | -0.1094 |
| 101 | $C_{cyc}$-OH | -0.7038 | **** | -0.8063 | **** | -0.6692 | **** | -0.6998 | **** | -0.5360 | **** | -0.5070 | **** | -0.7038 |
| 102 | $>N_{cyc}$-$CH_3$ | -1.2066 | 0.1103 | -0.2019 | 0.2099 | -0.4868 | 0.1756 | 0.0853 | 0.6472 | 0.0919 | 0.0726 | -0.2492 | 0.1523 | -1.2066 |
| 103 | $>N_{cyc}$-$CH_2$ | -0.7739 | 0.3840 | -0.6276 | 0.2840 | -0.7202 | 0.4209 | -1.8555 | 0.4905 | -0.5019 | 0.3356 | -0.6378 | 0.2710 | -0.7739 |
| 104 | AROMRINGs$^1$s$^2$ | -0.1769 | 0.1723 | -0.1816 | 0.2097 | -0.1361 | 0.1386 | -0.0783 | 0.1272 | -0.1280 | 0.1021 | -0.0969 | 0.1780 | -0.1769 |
| 105 | AROMRINGs$^1$s$^3$ | -0.1773 | -0.1755 | 0.1415 | - | -0.0428 | -0.2394 | -0.2547 | -0.3629 | -0.4406 | -0.1069 | -0.1728 | -0.1209 | -0.1773 |
| 106 | AROMRINGs$^1$s$^4$ | -0.0844 | -0.0608 | -0.1789 | - | -0.1036 | 0.0204 | -0.0250 | 0.0392 | -0.1419 | 0.0602 | -0.1504 | 0.1637 | -0.0844 |
| 107 | AROMRINGs$^1$s$^2$s$^3$ | -0.8970 | -0.2962 | -1.0224 | - | -1.3342 | -0.6484 | -1.2374 | -1.0101 | -1.3327 | -0.6523 | -1.2496 | -0.4397 | -0.8970 |
| 108 | AROMRINGs$^1$s$^2$s$^4$ | 0.0620 | 0.0797 | 0.0487 | - | 0.0410 | -0.1312 | 0.1380 | -0.1701 | -0.0209 | -0.1722 | -0.0483 | -0.1165 | 0.0620 |
| 109 | AROMRINGs$^1$s$^3$s$^5$ | 0.9750 | 0.1039 | 0.5147 | - | 0.3431 | 0.0309 | 0.4799 | -0.4711 | 0.9596 | -0.0603 | 0.9389 | -0.0419 | 0.9750 |
| 110 | AROMRINGs$^1$s$^2$s$^3$s$^4$ | -0.3576 | -0.1394 | -0.5347 | - | -0.6996 | -0.2504 | -0.2893 | -1.2009 | -0.4443 | 0.0538 | -0.3641 | -0.0909 | -0.3576 |
| 111 | AROMRINGs$^1$s$^2$s$^3$s$^5$ | 0.3356 | -0.0381 | 0.3463 | 0.0338 | 0.2977 | -0.0871 | 0.3619 | -0.2276 | 0.5856 | -0.0414 | 0.4694 | -0.0448 | 0.3356 |
| 112 | AROMRINGs$^1$s$^2$s$^4$s$^5$ | 0.4565 | 0.0981 | 0.2489 | 0.2706 | 0.0031 | 0.0371 | 0.1238 | 0.5343 | -0.1304 | -0.1554 | -0.0715 | 0.1909 | 0.4565 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 113 | PYRIDINEs$^{2}$ | -3.5683 | **** | -1.8976 | **** | -1.7913 | **** | -3.4187 | **** | -1.3125 | **** | -2.0497 | **** | -3.5683 |
| 114 | PYRIDINEs$^{3}$ | 0.6433 | **** | 0.4661 | **** | 0.3197 | **** | 0.6967 | **** | 0.5743 | **** | 0.4823 | **** | 0.6433 |
| 115 | PYRIDINEs$^{4}$ | 0.4128 | **** | 0.3197 | **** | -0.1549 | **** | 0.2061 | **** | 0.1359 | **** | -0.1956 | **** | 0.4128 |
| 116 | PYRIDINEs$^{2}$s$^{3}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 117 | PYRIDINEs$^{2}$s$^{4}$ | 0.5067 | **** | 0.6051 | **** | 0.5708 | **** | 0.9500 | **** | 0.4179 | **** | 0.4536 | **** | 0.5067 |
| 118 | PYRIDINEs$^{2}$s$^{5}$ | **** | -0.2465 | **** | - | **** | -0.0442 | **** | 0.5791 | **** | 0.0300 | **** | 0.2084 | **** |
| 119 | PYRIDINEs$^{2}$s$^{6}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 120 | PYRIDINEs$^{3}$s$^{4}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 121 | PYRIDINEs$^{3}$s$^{5}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 122 | PYRIDINEs$^{2}$s$^{3}$s$^{6}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 123 | (CH$^{n}$=CH$^{m}$)$^{cyc}$-COOH | 1.9335 | **** | 1.3822 | **** | 1.4197 | **** | 3.0471 | **** | 0.5406 | **** | 1.0354 | **** | 1.9335 |
| 124 | AROMRINGs$^{1}$s$^{2}$s$^{3}$s$^{4}$s$^{5}$ | -1.4672 | -0.2969 | -1.5435 | 0.0898 | -1.1716 | 0.4501 | -1.2833 | 0.7438 | -3.0130 | 0.3279 | -2.1132 | -0.1988 | -1.4672 |
| 125 | aC-NHCOCH$_2$N | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 126 | (N=C)$_{cyc}$-CH$_3$ | 1.4459 | 0.0987 | 0.6505 | 0.2277 | 0.8497 | 0.2606 | 0.8816 | -0.3916 | 0.3579 | 0.4891 | 0.5249 | 0.5919 | 1.4459 |
| 127 | aC-CONH(CH$_2$)$_2$N | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 128 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | 1.0075 | -0.0458 | 0.8867 | - | 0.7118 | 0.0687 | 0.5428 | 0.1131 | 0.7967 | 0.0251 | 0.9353 | -0.0841 | 1.0075 |
| 129 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | 0.7076 | **** | -0.3532 | **** | -0.1085 | **** | -0.3788 | **** | -0.6173 | **** | -0.6392 | **** | 0.7076 |
| 130 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | 0.5896 | **** | 0.7720 | **** | 0.8423 | **** | 0.5188 | **** | 1.1647 | **** | 1.2373 | **** | 0.5896 |

[a] The symbols $EUA_{C\ 2j}$, $EUA_{NC\ 2j}$, $ERA_{C\ 2j}$, $ERA_{NC\ 2j}$, $EFW_{C\ 2j}$, $EFW_{NC\ 2j}$, $ESW_{C\ 2j}$, $ESW_{NC\ 2j}$, $ENS_{C\ 2j}$, $ENS_{NC\ 2j}$, $EAS_{C\ 2j}$, and $EAS_{NC\ 2j}$ represent the contributions ($D_j$) of the second-order groups for the corresponding properties.

**Table S4. MG Method Based Property Models Analysed Using Step-Wise Regression Method: Third-Order Groups and their Contributions [a] for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), and PCO**

| | Group | $LC50(FM)_{3k}$ | $LC50(DM)_{3k}$ | $LD50_{3k}$ | $LogWs_{3k}$ | $BCF_{3k}$ | $PEL_{3k}$ | $PCO_{3k}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | HOOC-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | 0.0498 | **** | -0.4908 | 0.1790 | **** | -0.4810 | -0.0249 |
| 2 | $NH_n$-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | **** | **** | **** | 1.0484 | **** | **** | **** |
| 3 | $NH_2$-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | -0.2787 | **** | **** | **** | **** |
| 4 | OH-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | 0.1759 | 1.3600 | **** | **** | -0.1240 |
| 5 | OH-$(CH_p)_k$-O-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 6 | OH-$(CH_p)_k$-S-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 7 | OH-$(CH_p)_k$-$NH_x$-$(CH_n)_m$-OH ($m,k$>0; $p,n,x$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 8 | $CH_p$-O-$(CH_n)_m$-OH ($m$>2; $n,p$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 9 | $NH_2$-$(CH_n)_m$-$NH_2$ ($m$>2; $n$ in 0..2)) | -0.6399 | **** | 0.3552 | **** | **** | 0.5413 | **** |
| 10 | $NH_k$-$(CH_n)_m$-$NH_2$ ($m$>2; $k$ in 0..1; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 11 | SH-$(CH_n)_m$-SH ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 12 | NC-$(CH_n)_m$-CN ($m$>2) | -0.7496 | **** | 1.1305 | 1.0800 | **** | -0.1657 | **** |
| 13 | COO-$(CH_n)_m$-OOC ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 14 | aC-$(CH_n$=$CH_m)_{cyc}$ (fused rings) ($n,m$ in 0..1) | -0.1522 | -0.3984 | 0.1430 | 0.6020 | -0.1148 | -0.8684 | 0.5575 |
| 15 | aC-aC (different rings) | 0.1281 | 0.2825 | 0.0735 | -0.2230 | 0.0330 | 0.0122 | **** |
| 16 | aC-$CH_{ncyc}$ (different rings) ($n$ in 0..1) | -0.1468 | 0.0000 | 0.1622 | 0.5020 | -0.1038 | **** | **** |
| 17 | aC-$CH_{ncyc}$ (fused rings) ($n$ in 0..1) | -0.2569 | -0.0249 | 0.0701 | 0.3690 | -0.0008 | 0.5251 | -0.0309 |
| 18 | aC-$(CH_n)_m$-aC (different rings) ($m$>1; $n$ in 0..2) | 0.0710 | **** | -0.4276 | -1.3000 | -2.5311 | **** | **** |
| 19 | aC-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | 0.9803 | 0.0000 | -0.0149 | -0.3336 | **** | **** | **** |
| 20 | $CH_{cyc}$-$CH_{cyc}$ (different rings) | **** | **** | -0.2107 | 0.6122 | **** | **** | 0.2449 |
| 21 | $CH_{cyc}$-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 22 | $CH_{multiring}$ | 0.0499 | -0.0154 | -0.0344 | -0.0247 | 0.2805 | 0.1166 | 0.0582 |
| 23 | $C_{multiring}$ | -0.0662 | -0.0052 | 0.0023 | 0.0710 | -0.3160 | -0.2036 | -0.0162 |
| 24 | aC-$CH_m$-aC (different rings) ($m$ in 0..2) | -0.2276 | -0.4233 | -0.0873 | 0.2310 | 0.0506 | -0.9455 | 0.0509 |
| 25 | aC-$(CH_m$=$CH_n)$-aC (different rings) ($m,n$ in 0..2) | **** | -0.5224 | -0.3410 | -0.8542 | -0.0416 | **** | **** |
| 26 | $(CH_m$=$C)_{cyc}$-CH=CH-$(C$=$CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** |
| 27 | $(CH_m$=$C)_{cyc}$-$CH_p$-$(C$=$CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** |
| 28 | aC-CO-aC (different rings) | -0.2159 | **** | 0.1961 | 1.0600 | -0.4738 | **** | **** |
| 29 | aC-$CH_m$-CO-aC (different rings) ($m$ in 0..2) | **** | **** | -0.3475 | -0.5809 | 0.5000 | **** | **** |
| 30 | aC-CO-$(C$=$CH_n)_{cyc}$ (different rings) ($n$ in 0..1) | **** | **** | -0.0519 | -0.4667 | **** | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 31 | aC-CO-CO-aC (different rings) | **** | **** | **** | 0.1407 | **** | **** | **** |
| 32 | aC-$CO_{cyc}$ (fused rings) | 0.2092 | 0.2980 | -0.0138 | 0.0380 | -0.0285 | -1.1411 | -0.1487 |
| 33 | aC-CO-$(CH_n)_m$-CO-aC (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 34 | aC-CO-$CH_{ncyc}$ (different rings) ($n$ in 0..1) | **** | **** | **** | -0.6102 | **** | **** | **** |
| 35 | aC-CO-$NH_n$-aC (different rings) ($n$ in 0..1) | -0.0607 | **** | -0.2718 | -0.5870 | -0.2114 | **** | **** |
| 36 | aC-$NH_nCONH_m$-aC (different rings) ($n,m$ in 0..1) | 1.4228 | -0.0895 | **** | 0.8040 | -0.6021 | **** | **** |
| 37 | aC-CO-$N_{cyc}$ (different rings) | **** | **** | 0.3225 | -1.0070 | **** | **** | **** |
| 38 | aC-$S_{cyc}$ (fused rings) | -0.8909 | -0.1716 | 0.0828 | 0.1650 | -0.0918 | -0.5763 | **** |
| 39 | aC-S-aC (different rings) | -1.9636 | **** | -0.5044 | 0.4514 | -1.7418 | **** | **** |
| 40 | aC-$PO_n$-aC (different rings) ($n$ in 0..4) | **** | **** | **** | **** | **** | **** | **** |
| 41 | aC-$SO_n$-aC (different rings) ($n$ in 1..4) | **** | 0.1267 | 0.0683 | -0.1410 | **** | **** | **** |
| 42 | aC-$NH_{ncyc}$ (fused rings) ($n$ in 0..1) | 0.2702 | 0.9450 | -0.0298 | -0.1200 | 0.0486 | 0.5000 | **** |
| 43 | aC-NH-aC (different rings) | 0.2900 | **** | -0.0049 | -0.1120 | 0.2117 | -0.5446 | **** |
| 44 | aC-$(C=N)_{cyc}$ (different rings) | **** | **** | -0.1266 | 0.5247 | -0.3031 | **** | **** |
| 45 | aC-$(N=CH_n)_{cyc}$ (fused rings) ($n$ in 0..1) | -0.4338 | -0.5251 | 0.0640 | -0.0249 | -0.1578 | **** | **** |
| 46 | aC-$(CH_n=N)_{cyc}$ (fused rings) ($n$ in 0..1) | **** | **** | -0.1276 | 0.4628 | **** | **** | **** |
| 47 | aC-O-$CH_n$-aC (different rings) ($n$ in 0..2) | **** | **** | -0.2231 | 0.8493 | **** | **** | **** |
| 48 | aC-O-aC (different rings) | 0.1437 | 0.0020 | -0.0688 | -0.3224 | -0.0149 | -0.1030 | **** |
| 49 | aC-$CH_n$-O-$CH_m$-aC (different rings) ($n,m$ in 0..2) | **** | **** | -0.7009 | 0.2160 | **** | **** | **** |
| 50 | aC-$O_{cyc}$ (fused rings) | -0.4414 | -0.5119 | 0.0329 | -0.3410 | -0.0203 | **** | **** |
| 51 | AROM.FUSED[2] | 0.0032 | 0.1720 | -0.1119 | -0.0962 | 0.0643 | 0.0763 | -0.0183 |
| 52 | AROM.FUSED[2]$s^1$ | 0.1792 | 0.4007 | 0.2156 | 0.1030 | 0.2267 | -0.1548 | -0.0209 |
| 53 | AROM.FUSED[2]$s^2$ | 0.2418 | -0.1498 | -0.0803 | -0.0515 | -0.0089 | 1.8721 | 0.0295 |
| 54 | AROM.FUSED[2]$s^2s^3$ | -0.2860 | 1.2197 | 0.3000 | 0.1132 | 0.2159 | **** | 0.1743 |
| 55 | AROM.FUSED[2]$s^1s^4$ | -0.2961 | **** | -0.2363 | -0.2116 | -0.0322 | **** | **** |
| 56 | AROM.FUSED[2]$s^1s^2$ | **** | **** | 0.1553 | 0.0217 | -0.4516 | **** | **** |
| 57 | AROM.FUSED[2]$s^1s^3$ | **** | **** | 0.3250 | -0.0558 | 0.2075 | **** | **** |
| 58 | AROM.FUSED[3] | 0.0955 | -0.3703 | -0.1814 | -0.2256 | 0.0443 | 0.1575 | **** |
| 59 | AROM.FUSED[4a] | **** | **** | -0.2323 | -0.7000 | -0.0350 | -0.1471 | **** |
| 60 | AROM.FUSED[4a]$s^1$ | **** | **** | -0.4303 | 0.3143 | -0.0103 | **** | **** |
| 61 | AROM.FUSED[4a]$s^1s^4$ | **** | **** | 1.1705 | 0.3697 | **** | **** | **** |
| 62 | AROM.FUSED[4p] | 1.5851 | -0.0972 | -0.1580 | 0.0143 | -0.1692 | -0.0713 | **** |
| 63 | AROM.FUSED[4p]$s^3s^4$ | **** | **** | 0.2203 | 1.8369 | **** | **** | **** |
| 64 | PYRIDINE.FUSED[2] | -0.0274 | -0.4928 | 0.0166 | 0.4989 | -1.4055 | **** | **** |
| 65 | PYRIDINE.FUSED[2-iso] | **** | -0.3098 | 0.6246 | 1.3160 | **** | **** | **** |
| 66 | PYRIDINE.FUSED[4] | 0.7501 | -0.0203 | 0.1412 | 1.9800 | -0.3701 | -0.3792 | **** |
| 67 | aC-N-$CH_{cyc}$ (different rings) | **** | **** | 0.1538 | -0.1524 | **** | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 68 | N multiring | -0.0136 | 0.6590 | -0.0030 | 0.1349 | **** | **** | 0.0258 |
| 69 | $N_{cyc}$-$(CH_2)_3$-$N_{cyc}$ (different rings) | **** | **** | -0.0732 | -0.1746 | **** | **** | **** |
| 70 | aC-$COCH_2CH_2$-aC (different rings) | **** | **** | -0.3699 | -0.6448 | **** | **** | **** |
| 71 | aC-O-$(CH_2)_2$-$N_{cyc}$ (different rings) | **** | **** | -0.5047 | **** | **** | **** | **** |
| 72 | aC-$CH(OH)(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** |
| 73 | Ncyc-$(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | -0.1872 | **** | **** | **** | **** |
| 74 | aC-$CONHCH_2$-$CH_{cyc}$ (different rings) | **** | **** | 0.5619 | **** | **** | **** | **** |

[a] The symbols $LC50(FM)_{3k}$, $LC50(DM)_{3k}$, $LD50_{3k}$, $LogWs_{3k}$, $BCF_{3k}$, $PEL_{3k}$, $PCO_{3k}$ represent the contributions ($E_k$) of the third-order groups for the corresponding properties. Note that there are no third-order contributions for GWP, ODP and AP since these properties belong to chemicals whose structure is defined by first-order groups alone.

**Table S4 (continued). MG Method Based Property Models Analysed Using Step-Wise Regression Method: Third-Order Groups and their Contributions [a] for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| | Group | $EUA_{C\ 3k}$ | $EUA_{NC3k}$ | $ERA_{C\ 3k}$ | $ERA_{NC3k}$ | $EFW_{C\ 3k}$ | $EFW_{NC\ 3k}$ | $ESW_{C\ 2j}$ | $ESW_{NC\ 3k}$ | $ENS_{C\ 3k}$ | $ENS_{NC\ 3k}$ | $EAS_{C\ 3k}$ | $EAS_{NC\ 3k}$ | $EUA_{C\ 3k}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HOOC-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 2 | $NH_n$-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 3 | $NH_2$-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 4 | OH-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 5 | OH-$(CH_p)_k$-O-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 6 | OH-$(CH_p)_k$-S-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 7 | OH-$(CH_p)_k$-$NH_x$-$(CH_n)_m$-OH ($m,k$>0; $p,n,x$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 8 | $CH_p$-O-$(CH_n)_m$-OH ($m$>2; $n,p$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 9 | $NH_2$-$(CH_n)_m$-$NH_2$ ($m$>2; $n$ in 0..2)) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 10 | $NH_k$-$(CH_n)_m$-$NH_2$ ($m$>2; $k$ in 0..1; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | SH-$(CH_n)_m$-SH ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 12 | NC-$(CH_n)_m$-CN ($m$>2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | COO-$(CH_n)_m$-OOC ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 14 | aC-$(CH_n{=}CH_m)_{cyc}$ (fused rings) ($n,m$ in 0..1) | -0.1215 | **** | -0.1473 | **** | 0.1250 | **** | -0.0685 | **** | 0.2146 | **** | 0.2365 | **** | -0.1215 |
| 15 | aC-aC (different rings) | -0.2468 | 0.0671 | -0.1842 | -0.2102 | -0.2456 | -0.0737 | -0.3819 | -0.1675 | -0.1974 | -0.0488 | -0.0378 | -0.0141 | -0.2468 |
| 16 | aC-$CH_{ncyc}$ (different rings) ($n$ in 0..1) | 0.2361 | 0.3257 | 0.1430 | 0.3934 | 0.3252 | 0.4016 | 0.2774 | 0.7521 | -0.0755 | 0.1389 | -0.1467 | 0.0749 | 0.2361 |
| 17 | aC-$CH_{ncyc}$ (fused rings) ($n$ in 0..1) | 0.1459 | 0.2401 | -0.0771 | 0.5147 | 0.0586 | 0.3469 | 0.2672 | 0.2325 | -0.1877 | 0.2791 | -0.1218 | 0.3119 | 0.1459 |
| 18 | aC-$(CH_n)_m$-aC (different rings) ($m$>1; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 19 | aC-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 20 | $CH_{cyc}$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 21 | $CH_{cyc}$-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 22 | $CH_{multiring}$ | 0.0300 | 0.0278 | 0.0391 | -0.0003 | -0.0363 | -0.0292 | -0.0287 | -0.0117 | -0.0504 | -0.0425 | -0.0394 | -0.0300 | 0.0300 |
| 23 | $C_{multiring}$ | -0.0165 | -0.0191 | -0.0075 | -0.0090 | 0.0119 | 0.0001 | 0.0147 | -0.0020 | 0.0377 | 0.0147 | 0.0267 | 0.0038 | -0.0165 |
| 24 | aC-$CH_m$-aC (different rings) ($m$ in 0..2) | -0.3033 | 0.0545 | -0.2308 | 0.0585 | -0.1917 | 0.0840 | -0.3588 | 0.0777 | 0.1200 | 0.1306 | 0.0679 | 0.1377 | -0.3033 |
| 25 | aC-$(CH_m{=}CH_n)$-aC (different rings) ($m,n$ in 0..2) | -1.2865 | **** | -1.5110 | **** | -0.9390 | **** | -1.3848 | **** | -0.2259 | **** | -0.3749 | **** | -1.2865 |
| 26 | $(CH_m{=}C)_{cyc}$-CH=CH-$(C{=}CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 27 | $(CH_m{=}C)_{cyc}$-$CH_p$-$(C{=}CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 28 | aC-CO-aC (different rings) | 0.3796 | **** | 0.7024 | **** | 0.7866 | **** | 0.5009 | **** | 0.7473 | **** | 0.7154 | **** | 0.3796 |
| 29 | aC-$CH_m$-CO-aC (different rings) ($m$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | aC-CO-(C=CH$_n$)$_{cyc}$ (different rings) ($n$ in 0..1) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 31 | aC-CO-CO-aC (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 32 | aC-CO$_{cyc}$ (fused rings) | -0.0933 | 0.0876 | -0.0655 | 0.0969 | -0.0074 | 0.1370 | -0.1558 | 0.2045 | -0.0724 | -0.1393 | -0.0848 | -0.1407 | -0.0933 |
| 33 | aC-CO-(CH$_n$)$_m$-CO-aC (different rings) ($m$>0; $n$ in | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 34 | aC-CO-CH$_{ncyc}$ (different rings) ($n$ in 0..1) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 35 | aC-CO-NH$_n$-aC (different rings) ($n$ in 0..1) | **** | -0.1635 | **** | 0.0182 | **** | -0.1253 | **** | -0.0293 | **** | -0.2030 | **** | -0.2010 | **** |
| 36 | aC-NH$_n$CONH$_m$-aC (different rings) ($n$,$m$ in 0..1) | 0.0224 | **** | 0.1302 | **** | 0.0626 | **** | -0.1130 | **** | 0.1627 | **** | 0.1987 | **** | 0.0224 |
| 37 | aC-CO-N$_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 38 | aC-S$_{cyc}$ (fused rings) | 1.0646 | 0.5747 | 0.9674 | 1.0012 | 1.0749 | 0.8358 | 1.1511 | 0.4969 | 1.7345 | 0.1996 | 1.7415 | 0.4978 | 1.0646 |
| 39 | aC-S-aC (different rings) | 0.1689 | **** | 0.3578 | **** | 0.2071 | **** | 0.0500 | **** | 0.2837 | **** | 0.3008 | **** | 0.1689 |
| 40 | aC-PO$_n$-aC (different rings) ($n$ in 0..4) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 41 | aC-SO$_n$-aC (different rings) ($n$ in 1..4) | -0.6611 | **** | -0.3872 | **** | -0.2601 | **** | -0.1723 | **** | -0.2444 | **** | -0.4162 | **** | -0.6611 |
| 42 | aC-NH$_{ncyc}$ (fused rings) ($n$ in 0..1) | 0.2286 | 0.5134 | 0.0775 | 0.6853 | 0.1433 | 0.5197 | 0.3180 | 0.2492 | 0.3238 | 0.4479 | 0.2342 | 0.6158 | 0.2286 |
| 43 | aC-NH-aC (different rings) | 3.0372 | 0.4637 | 2.9764 | 0.9832 | 2.6222 | 0.2142 | 2.1095 | 0.2696 | 2.7532 | 0.3378 | 2.6061 | 0.2319 | 3.0372 |
| 44 | aC-(C=N)$_{cyc}$ (different rings) | 0.3105 | 0.5157 | 0.0364 | 0.3922 | -0.3900 | 0.4152 | 0.3847 | 0.0866 | -1.1713 | -0.3841 | -0.9995 | -0.0549 | 0.3105 |
| 45 | aC-(N=CH$_n$)$_{cyc}$ (fused rings) ($n$ in 0..1) | -0.3841 | -0.2992 | -0.3342 | -0.2709 | -0.2907 | -0.3074 | 0.1324 | 0.0130 | -0.5998 | -0.3192 | -0.4553 | -0.5872 | -0.3841 |
| 46 | aC-(CH$_n$=N)$_{cyc}$ (fused rings) ($n$ in 0..1) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 47 | aC-O-CH$_n$-aC (different rings) ($n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 48 | aC-O-aC (different rings) | 0.6499 | -0.0274 | 0.4143 | -0.0110 | 0.4380 | -0.0296 | 0.5488 | -0.0890 | 1.1196 | 0.0980 | 0.9371 | 0.0870 | 0.6499 |
| 49 | aC-CH$_n$-O-CH$_m$-aC (different rings) ($n$,$m$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 50 | aC-O$_{cyc}$ (fused rings) | 0.0383 | -0.6411 | 0.0370 | -0.8716 | -0.1231 | -0.8194 | 0.0243 | -0.9153 | -0.0054 | -0.4711 | -0.0885 | -0.4723 | 0.0383 |
| 51 | AROM.FUSED[2] | 0.1136 | -0.2310 | 0.2166 | -0.4164 | 0.0549 | -0.3894 | 0.1091 | -0.3073 | 0.0170 | -0.1885 | 0.0079 | -0.1675 | 0.1136 |
| 52 | AROM.FUSED[2]s$^1$ | 0.1925 | 0.5089 | 0.0158 | 0.6586 | -0.1485 | 0.8062 | 0.0589 | 0.6587 | 0.2571 | 0.3565 | 0.3012 | 0.5469 | 0.1925 |
| 53 | AROM.FUSED[2]s$^2$ | -0.2109 | 0.3013 | -0.1850 | 0.6488 | -0.0357 | 0.3817 | -0.2511 | 0.4418 | 0.3028 | 0.1445 | 0.2705 | -0.0524 | -0.2109 |
| 54 | AROM.FUSED[2]s$^2$s$^3$ | -1.5896 | 1.2673 | 0.6844 | 1.8541 | 0.7961 | 1.6729 | -1.6291 | 2.2637 | -1.0883 | 0.7740 | -0.9250 | 0.6817 | -1.5896 |
| 55 | AROM.FUSED[2]s$^1$s$^4$ | -0.0973 | **** | 0.0836 | **** | -0.2053 | **** | -0.0020 | **** | -0.0769 | **** | -0.0043 | **** | -0.0973 |
| 56 | AROM.FUSED[2]s$^1$s$^2$ | 0.7666 | -1.2625 | -0.2823 | -0.6895 | -0.3718 | -0.0288 | 0.8986 | -1.2125 | 0.9963 | 2.6709 | 0.9649 | 1.8788 | 0.7666 |
| 57 | AROM.FUSED[2]s$^1$s$^3$ | -0.6426 | **** | -0.8108 | **** | -0.6896 | **** | -0.8285 | **** | -1.0441 | **** | -0.9791 | **** | -0.6426 |
| 58 | AROM.FUSED[3] | -0.1308 | 0.1254 | 0.3385 | 0.1587 | 0.1683 | -0.0901 | -0.3007 | -0.0059 | 0.2048 | -0.1068 | 0.3250 | -0.1080 | -0.1308 |
| 59 | AROM.FUSED[4a] | 0.3064 | **** | 0.7653 | 1.1669 | 0.5645 | 1.1140 | 0.5066 | 1.0208 | 0.1222 | 0.5681 | 0.6221 | 0.5363 | 0.3064 |
| 60 | AROM.FUSED[4a]s$^1$ | 2.1227 | **** | 1.8670 | **** | 1.8023 | **** | 2.4139 | **** | 1.6935 | **** | 1.9882 | **** | 2.1227 |
| 61 | AROM.FUSED[4a]s$^1$s$^4$ | -1.0810 | **** | -2.0746 | **** | -1.9895 | **** | -2.5320 | **** | -1.6361 | **** | -1.3919 | **** | -1.0810 |
| 62 | AROM.FUSED[4p] | -0.0590 | -0.1254 | -0.2886 | -0.4931 | -0.0769 | -0.0628 | 0.0270 | -0.1591 | -0.0097 | 0.0760 | -0.1541 | 0.0488 | -0.0590 |
| 63 | AROM.FUSED[4p]s$^3$s$^4$ | 0.2168 | **** | 0.1719 | **** | -0.0581 | **** | 0.5679 | **** | -0.1928 | **** | -0.0284 | **** | 0.2168 |
| 64 | PYRIDINE.FUSED[2] | -0.7523 | **** | -0.5874 | **** | -0.5573 | **** | 0.2154 | **** | -0.4242 | **** | -0.2953 | **** | -0.7523 |
| 65 | PYRIDINE.FUSED[2-iso] | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 66 | PYRIDINE.FUSED[4] | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 67 | aC-N-$CH_{cyc}$ (different rings) | **** | **** | 0.0000 | **** | 0.0000 | **** | **** | **** | 0.0000 | **** | 0.0000 | **** | **** |
| 68 | N multiring | -0.2136 | -0.4632 | -0.0115 | -0.4748 | -0.0907 | -0.0129 | -0.2649 | 0.0949 | 0.1129 | -0.0060 | 0.1290 | 0.0151 | -0.2136 |
| 69 | $N_{cyc}$-$(CH_2)_3$-$N_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 70 | aC-$COCH_2CH_2$-aC (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 71 | aC-O-$(CH_2)_2$-$N_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 72 | aC-$CH(OH)(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 73 | Ncyc-$(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 74 | aC-$CONHCH_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

[a] The symbols $EUA_{C\ 3k}$, $EUA_{NC\ 3k}$, $ERA_{C\ 3k}$, $ERA_{NC\ 3k}$, $EFW_{C\ 3k}$, $EFW_{NC\ 3k}$, $ESW_{C\ 3k}$, $ESW_{NC\ 3k}$, $ENS_{C\ 3k}$, $ENS_{NC\ 3k}$, $EAS_{C\ 3k}$, and $EAS_{NC\ 3k}$ represent the contributions ($E_k$) of the third-order groups for the corresponding properties.

**Table S5. MG Method Based Property Models Analysed Using <u>Simultaneous Regression Method</u>: First-Order Groups and their Contributions [a] for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$,, BCF, PEL(OSHA-TWA), PCO, GWP, ODP, and AP.**

| | Group | $LC50(FM)_{1i}$ | $LC50(DM)_{1i}$ | $LD50_{1i}$ | $LogWs_{1i}$ | $BCF_{1i}$ | $PEL_{1i}$ | $PCO_{1i}$ | $GWP_{1i}$ | $ODP_{1i}$ | $AP_{1i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $CH_3$ | 0.0556 | -0.3263 | -0.0620 | -5.1279 | 0.6703 | 0.7059 | 0.1421 | 0.3880 | -0.9453 | -0.1290 |
| 2 | $CH_2$ | 0.3058 | 0.1806 | 0.0206 | -5.0803 | 0.0883 | 0.0742 | 0.0452 | -1.0699 | **** | **** |
| 3 | CH | 0.2640 | -0.3385 | 0.0834 | -4.7909 | -0.6251 | -0.5252 | -0.1351 | **** | **** | **** |
| 4 | C | -0.5591 | -2.3767 | 0.1524 | -4.9670 | -1.2271 | -1.1333 | -0.2452 | **** | **** | **** |
| 5 | $CH_2$=CH | 1.1915 | -4.7284 | 0.1567 | -9.5872 | 0.6777 | 2.4164 | -0.3641 | **** | **** | **** |
| 6 | CH=CH | 0.3986 | -1.6245 | 0.1181 | -9.5886 | **** | 1.3391 | -0.7034 | **** | **** | **** |
| 7 | $CH_2$=C | 0.7062 | -0.9381 | 0.1993 | -9.5121 | 0.3400 | 1.1805 | -0.5276 | **** | **** | **** |
| 8 | CH=C | 0.9899 | 0.0393 | 0.3274 | -9.4355 | -0.3654 | 0.8654 | -0.8041 | **** | **** | **** |
| 9 | C=C | 1.2892 | -0.9820 | 0.2934 | -9.5320 | 0.5447 | 1.2985 | -0.5092 | **** | **** | **** |
| 10 | $CH_2$=C=CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | $CH_2$=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 12 | C=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | CH≡C | 0.9001 | **** | 0.0737 | -8.7940 | **** | 3.0437 | -0.2592 | **** | **** | **** |
| 14 | C≡C | 1.5163 | 3.6012 | 0.3346 | -8.7290 | **** | **** | -0.9453 | **** | **** | **** |
| 15 | aCH | 0.1574 | -0.0680 | 0.0200 | -4.5565 | 0.2711 | 0.3182 | 0.0263 | **** | **** | **** |
| 16 | aC fused with aromatic ring | -0.4620 | -0.7792 | 0.0729 | -4.7557 | 0.0144 | 0.2181 | -0.0470 | **** | **** | **** |
| 17 | aC fused with non-aromatic ring | 0.0006 | -0.7416 | 0.0358 | -4.4965 | -0.0080 | 0.0889 | -0.0958 | **** | **** | **** |
| 18 | aC except as above | 0.0893 | -0.4944 | 0.0331 | -4.4476 | -0.2361 | 0.3662 | 0.8361 | **** | **** | **** |
| 19 | aN in aromatic ring | 0.1432 | -0.3506 | 0.0384 | -4.4675 | 0.1866 | 2.1309 | **** | **** | **** | **** |
| 20 | aC-$CH_3$ | 0.3434 | 0.3598 | 0.0616 | -9.5780 | 0.3631 | 0.8824 | -0.0776 | **** | **** | **** |
| 21 | aC-$CH_2$ | 0.7309 | 0.1167 | 0.1132 | -9.5168 | -0.0324 | -0.1488 | -0.1815 | **** | **** | **** |
| 22 | aC-CH | 0.6725 | 2.8366 | 0.5066 | -9.3337 | -0.9742 | 0.8149 | -0.2730 | **** | **** | **** |
| 23 | aC-C | 0.1056 | 1.9398 | 0.5653 | -9.8828 | -0.8256 | -5.0862 | -0.4485 | **** | **** | **** |
| 24 | aC-CH=$CH_2$ | 0.7720 | 0.1539 | 0.0850 | -14.1770 | 1.2395 | 0.9606 | 0.5094 | **** | **** | **** |
| 25 | aC-CH=CH | -0.3767 | **** | 0.1680 | -14.3619 | 0.2045 | **** | 1.0404 | **** | **** | **** |
| 26 | aC-C=$CH_2$ | **** | **** | -0.7064 | -14.4125 | -0.2210 | 0.2660 | 0.1505 | **** | **** | **** |
| 27 | aC-C≡CH | **** | **** | **** | -13.6327 | **** | **** | **** | **** | **** | **** |
| 28 | aC-C≡C | **** | **** | -0.4446 | -14.2762 | **** | **** | **** | **** | **** | **** |
| 29 | OH | -0.4918 | -1.9704 | -0.2449 | -4.9955 | 0.1957 | 1.3297 | 0.0674 | **** | **** | **** |
| 30 | aC-OH | 0.2408 | 0.1957 | 0.0580 | -9.6506 | -0.4151 | 1.5157 | 0.4048 | **** | -1.5842 | -0.0769 |
| 31 | COOH | -0.6851 | -0.7502 | 0.0011 | -14.6373 | -1.3264 | 2.5364 | 0.1482 | **** | **** | **** |
| 32 | aC-COOH | -0.0709 | -1.9953 | 0.0388 | -19.3390 | -1.0929 | 1.4314 | **** | **** | **** | **** |
| 33 | $CH_3CO$ | -0.0161 | -2.5920 | -0.0697 | -14.3754 | 0.2364 | 0.8681 | 0.5458 | **** | **** | **** |
| 34 | $CH_2CO$ | -0.5132 | **** | 0.1381 | -14.4832 | -1.0508 | 0.9384 | -0.0817 | **** | **** | **** |
| 35 | CHCO | **** | **** | 0.4061 | -14.9310 | **** | **** | -0.0316 | **** | **** | **** |
| 36 | CCO | **** | **** | 1.1095 | -14.3234 | **** | -1.7554 | **** | **** | **** | **** |
| 37 | aC-CO | 0.4173 | 1.4830 | 0.1142 | -14.0458 | -0.9617 | 0.9302 | **** | **** | **** | **** |
| 38 | CHO | 0.3890 | -2.3399 | -0.1130 | -9.0480 | -0.4066 | 2.3638 | -0.2229 | **** | **** | **** |
| 39 | aC-CHO | 0.7570 | 0.7430 | -0.0527 | -14.3857 | -1.2851 | **** | **** | **** | **** | **** |
| 40 | $CH_3COO$ | 0.4260 | -1.3681 | -0.1990 | -19.6379 | -0.0072 | 1.1646 | 0.4531 | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | $CH_2COO$ | 0.2046 | 0.7330 | -0.0418 | -19.6496 | 0.0120 | 1.7849 | 0.1935 | **** | **** | **** |
| 42 | CHCOO | -0.4731 | **** | 0.1131 | -19.3341 | **** | **** | 0.1831 | **** | **** | **** |
| 43 | CCOO | -1.9971 | **** | 0.2846 | -19.4530 | **** | **** | **** | **** | **** | **** |
| 44 | HCOO | **** | **** | -0.2234 | -15.3161 | **** | 1.6436 | 0.8105 | **** | **** | **** |
| 45 | aC-COO | 0.3771 | 0.3883 | -0.0421 | -19.3085 | -1.1622 | 0.9798 | 0.0416 | **** | **** | **** |
| 46 | aC-OOCH | **** | **** | -0.2859 | **** | **** | **** | **** | **** | **** | **** |
| 47 | aC-OOC | 1.1756 | **** | 0.3537 | -19.5422 | -3.7123 | 1.6104 | **** | **** | **** | **** |
| 48 | COO except as above | 0.6212 | 6.6088 | 0.1822 | -14.5866 | -1.2944 | 0.1698 | -0.2389 | **** | **** | **** |
| 49 | $CH_3O$ | -0.3622 | 1.1271 | -0.0219 | -10.0648 | 0.2283 | 2.1170 | 0.1751 | 0.1245 | **** | **** |
| 50 | $CH_2O$ | -0.2490 | 0.1565 | 0.1148 | -9.8886 | -0.2561 | 0.9385 | -0.1353 | **** | **** | **** |
| 51 | CH-O | -0.5300 | **** | 0.4559 | -9.9280 | -0.1236 | -0.6058 | -0.2934 | -1.8521 | **** | **** |
| 52 | C-O | **** | **** | 0.0697 | -9.7047 | -0.8483 | **** | **** | **** | **** | **** |
| 53 | aC-O | 0.0233 | 0.7468 | 0.1589 | -9.6786 | -0.2179 | 0.8384 | -0.5301 | **** | **** | **** |
| 54 | $CH_2NH_2$ | 0.0884 | 1.8188 | 0.0659 | -9.3586 | 0.3801 | 2.6815 | -0.3332 | **** | **** | **** |
| 55 | $CHNH_2$ | 0.0050 | 0.9501 | 0.4614 | -9.7816 | **** | 2.2834 | -0.6060 | **** | **** | **** |
| 56 | $CNH_2$ | 1.0340 | **** | 0.5006 | -8.8835 | **** | **** | **** | **** | **** | **** |
| 57 | $CH_3NH$ | -0.6909 | 0.1063 | -0.0037 | -9.4825 | **** | 2.8993 | -0.0143 | **** | **** | **** |
| 58 | $CH_2NH$ | 0.2411 | 0.1899 | 0.3476 | -9.3334 | -0.5586 | 1.5850 | 0.0159 | **** | **** | **** |
| 59 | CHNH | **** | -0.1355 | 0.2110 | -8.8354 | **** | 1.4112 | **** | **** | **** | **** |
| 60 | $CH_3N$ | 0.0656 | -0.2190 | 0.3385 | -9.5616 | -0.3696 | 1.2989 | -0.4572 | **** | **** | **** |
| 61 | $CH_2N$ | -0.9455 | -0.6699 | 0.4260 | -9.3848 | -1.4876 | 0.4199 | -0.4197 | **** | **** | **** |
| 62 | aC-$NH_2$ | -0.0787 | 1.0092 | 0.1302 | -9.8285 | -0.4375 | 2.4018 | **** | **** | **** | **** |
| 63 | aC-NH | 0.0385 | 1.0054 | 0.2339 | -9.6863 | -0.6826 | 1.4233 | **** | **** | **** | **** |
| 64 | aC-N | 0.0719 | 1.2340 | 0.3112 | -9.7588 | -0.7716 | 0.1870 | **** | **** | **** | **** |
| 65 | $NH_2$ except as above | 0.2122 | 0.3256 | 0.0820 | -5.0409 | -0.2637 | 2.0136 | -0.4508 | **** | **** | **** |
| 66 | CH=N | 2.7462 | 0.7708 | 0.1311 | -9.7606 | 0.3204 | **** | **** | **** | **** | **** |
| 67 | C=N | 1.5451 | 1.5013 | 0.2109 | -9.1601 | -1.3495 | **** | **** | **** | **** | **** |
| 68 | $CH_2CN$ | 0.2409 | **** | -0.0996 | -13.5065 | -0.0541 | 2.5521 | **** | **** | **** | **** |
| 69 | CHCN | 2.0033 | **** | 0.9700 | -13.6088 | -0.5524 | 2.0880 | **** | **** | **** | **** |
| 70 | CCN | 1.0562 | **** | 0.7209 | -12.9959 | **** | 0.9579 | **** | **** | **** | **** |
| 71 | aC-CN | 0.1265 | 0.3584 | 0.0653 | -13.5175 | -0.3325 | **** | **** | **** | **** | **** |
| 72 | CN except as above | -0.3055 | 0.6079 | 0.3035 | -8.9767 | 1.8953 | 2.0420 | 0.5544 | **** | **** | **** |
| 73 | $CH_2NCO$ | **** | **** | 0.0478 | **** | **** | 3.1925 | **** | **** | **** | **** |
| 74 | CHNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 75 | CNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 76 | aC-NCO | **** | **** | -0.1101 | **** | **** | 2.6914 | 0.0219 | **** | **** | **** |
| 77 | $CH2NO_2$ | **** | **** | 0.0869 | -20.1394 | **** | 1.9469 | 1.0619 | **** | **** | **** |
| 78 | $CHNO_2$ | **** | **** | 0.6395 | -20.4173 | -1.3892 | 1.5866 | **** | **** | **** | **** |
| 79 | $CNO_2$ | 1.1653 | **** | 1.1733 | -18.9795 | **** | -1.0086 | **** | **** | **** | **** |
| 80 | aC-$NO_2$ | 0.6761 | 0.1244 | 0.2759 | -20.2293 | -0.0575 | 2.0505 | 1.4104 | **** | **** | **** |
| 81 | $NO_2$ except as above | 1.0120 | 0.2969 | -0.0791 | -16.2802 | **** | 1.7993 | **** | **** | **** | -0.0775 |
| 82 | ONO | **** | **** | 0.8755 | **** | **** | **** | **** | **** | **** | **** |
| 83 | $ONO_2$ | -0.0709 | -0.0915 | 0.3217 | -21.3506 | **** | 1.9774 | **** | **** | **** | -0.0775 |
| 84 | $HCON(CH_2)_2$ | -0.2735 | **** | **** | -22.4833 | **** | **** | **** | **** | **** | **** |
| 85 | $HCONHCH_2$ | **** | **** | **** | -19.0479 | **** | **** | **** | **** | **** | **** |
| 86 | $CONH_2$ | -0.2099 | 3.7853 | 0.0938 | -14.4869 | **** | 2.9582 | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 87 | $CONHCH_3$ | 1.1894 | 0.3027 | 0.9970 | -19.3067 | -0.9446 | **** | **** | **** | **** | **** |
| 88 | $CONHCH_2$ | 1.0479 | -0.2683 | 0.1235 | -19.2945 | -1.7323 | **** | **** | **** | **** | **** |
| 89 | $CON(CH_3)_2$ | -0.7894 | **** | 0.6593 | -23.8705 | 2.1722 | 2.6902 | **** | **** | **** | **** |
| 90 | $CONCH_3CH_2$ | **** | **** | **** | -23.7820 | **** | **** | **** | **** | **** | **** |
| 91 | $CON(CH_2)_2$ | -0.5577 | -0.3976 | 0.3369 | -23.5732 | -1.5293 | **** | **** | **** | **** | **** |
| 92 | CONHCO | **** | **** | 0.6221 | -24.3563 | **** | **** | **** | **** | **** | **** |
| 93 | CONCO | 1.4698 | **** | 0.4604 | -23.7997 | **** | **** | **** | **** | **** | **** |
| 94 | $aC\text{-}CONH_2$ | -0.5257 | -1.5913 | 0.2189 | -19.1698 | **** | **** | **** | **** | **** | **** |
| 95 | aC-NH(CO)H | 0.0007 | **** | 0.2914 | -18.3036 | **** | **** | **** | **** | **** | **** |
| 96 | aC-N(CO)H | -0.0840 | **** | 0.2079 | **** | **** | **** | **** | **** | **** | **** |
| 97 | aC-CONH | 1.3217 | 2.6195 | 0.1560 | -18.7729 | -1.5536 | **** | **** | **** | **** | **** |
| 98 | aC-NHCO | -0.2182 | -0.1832 | 0.0072 | -18.9641 | -1.2849 | **** | **** | **** | **** | **** |
| 99 | aC-(N)CO | 0.0403 | 0.3436 | 0.1155 | -18.2987 | -2.3892 | **** | **** | **** | **** | **** |
| 100 | NHCONH | **** | **** | 0.1766 | -20.2163 | -2.8684 | **** | **** | **** | **** | **** |
| 101 | $NH_2CONH$ | **** | **** | -0.0339 | -20.0323 | **** | **** | **** | **** | **** | **** |
| 102 | $NH_2CON$ | **** | **** | 0.6960 | -19.7493 | **** | **** | **** | **** | **** | **** |
| 103 | NHCON | -1.1943 | **** | 0.6042 | -19.0795 | **** | **** | **** | **** | **** | **** |
| 104 | NCON | **** | -0.6828 | 0.4337 | -17.3382 | **** | **** | **** | **** | **** | **** |
| 105 | $aC\text{-}NHCONH_2$ | **** | **** | 0.0108 | -24.1940 | **** | **** | **** | **** | **** | **** |
| 106 | aC-NHCONH | -1.7196 | 1.9217 | 0.1090 | -25.0710 | -1.1863 | **** | **** | **** | **** | **** |
| 107 | NHCO except as above | 0.5780 | -1.3767 | 0.1661 | -14.2912 | 0.4049 | **** | **** | **** | **** | **** |
| 108 | $CH_2Cl$ | 0.6285 | -0.4994 | 0.3771 | -17.2028 | 0.3827 | 2.1156 | 0.3421 | -0.2757 | -1.4515 | **** |
| 109 | CHCl | 0.0499 | 0.5406 | 0.1587 | -17.0769 | 0.2987 | -0.6211 | 0.6224 | **** | -1.3742 | **** |
| 110 | CCl | **** | **** | 0.1362 | -17.1631 | **** | **** | **** | **** | **** | **** |
| 111 | $CHCl_2$ | 0.5537 | 0.0038 | 0.2677 | -28.9652 | 0.4192 | 1.5563 | 1.2588 | -0.0233 | -1.7873 | **** |
| 112 | $CCl_2$ | 1.2087 | **** | -0.0255 | -28.5311 | 1.0330 | **** | **** | **** | **** | **** |
| 113 | $CCl_3$ | 1.4243 | 0.8069 | 0.3228 | -41.4755 | 1.0662 | 1.6771 | 2.7478 | 1.8173 | -0.0399 | -0.0132 |
| 114 | $CH_2F$ | **** | **** | 2.1471 | -10.6675 | **** | **** | 1.8130 | 0.8584 | **** | **** |
| 115 | CHF | **** | **** | **** | -9.3840 | **** | **** | **** | -0.6900 | **** | **** |
| 116 | CF | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 117 | $CHF_2$ | 0.8937 | **** | 0.3330 | -17.3008 | **** | **** | 1.4605 | 1.6870 | -0.5628 | **** |
| 118 | $CF_2$ | 0.1228 | **** | 0.1107 | -18.6620 | 0.6414 | -0.5715 | **** | -0.0026 | -0.2492 | **** |
| 119 | $CF_3$ | 0.5746 | -0.4514 | 1.1768 | -23.8614 | 0.5289 | 1.2195 | 0.7594 | 2.1289 | 0.0883 | **** |
| 120 | $CCl_2F$ | **** | **** | -0.2444 | -35.8860 | 0.9305 | 0.3580 | 2.4605 | 2.1575 | 0.0131 | 0.1546 |
| 121 | HCClF | **** | **** | -0.5171 | -23.3119 | **** | -0.7500 | 1.9636 | 0.9439 | -1.4911 | **** |
| 122 | $CClF_2$ | **** | **** | 0.0285 | -29.8871 | 0.6942 | 0.5700 | 2.4605 | 2.0786 | -0.0409 | **** |
| 123 | aC-Cl | 0.6311 | 0.4746 | 0.1892 | -17.0037 | 0.5836 | 1.0971 | 0.6906 | **** | -0.2832 | -0.0290 |
| 124 | aC-F | 0.3314 | 0.0741 | 0.2803 | -10.8936 | 0.6116 | **** | 0.3160 | **** | **** | **** |
| 125 | aC-I | 0.9468 | **** | 0.2509 | -47.8151 | 0.7213 | **** | **** | **** | **** | **** |
| 126 | aC-Br | 0.8007 | 0.7128 | 0.4868 | -32.1342 | 0.4420 | **** | **** | **** | **** | **** |
| 127 | -I except as above | 0.9023 | -0.2683 | 0.4364 | -43.5393 | **** | 2.9991 | **** | -2.1289 | **** | **** |
| 128 | -Br except as above | 0.6833 | 0.2834 | 0.4352 | -27.4476 | 0.6304 | 1.4794 | 0.9350 | 1.0645 | 0.4320 | **** |
| 129 | -F except as above | 0.2140 | -0.3367 | 0.1492 | -6.7874 | 0.0237 | 1.9703 | 0.5273 | 1.9617 | -0.0074 | **** |
| 130 | -Cl except as above | 0.5444 | 0.0418 | -0.0421 | -12.3720 | 0.4939 | 2.3117 | 1.4774 | 1.2757 | -0.0056 | -0.0682 |
| 131 | CHNOH | 0.6580 | **** | 0.2879 | -15.3753 | **** | **** | **** | **** | **** | **** |
| 132 | CNOH | -0.3905 | **** | 0.1104 | -14.8064 | -1.2495 | **** | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 133 | aC-CHNOH | 1.3584 | **** | 0.9215 | **** | **** | **** | **** | **** | **** | **** |
| 134 | $OCH_2CH_2OH$ | -1.2180 | -1.8262 | -0.1391 | -19.4346 | -0.5730 | 1.7806 | 0.0004 | **** | **** | **** |
| 135 | $OCHCH_2OH$ | **** | **** | 0.0735 | -19.4893 | **** | -0.6850 | -0.1526 | **** | **** | **** |
| 136 | $OCH_2CHOH$ | 0.0290 | **** | -0.0840 | -19.6137 | **** | 0.8671 | -0.1808 | **** | **** | **** |
| 137 | -O-OH | **** | **** | -0.0096 | -9.9157 | **** | **** | 0.3052 | **** | **** | **** |
| 138 | $CH_2SH$ | 1.8331 | 2.4088 | 0.3595 | -16.8160 | -2.1222 | 3.4375 | **** | **** | **** | **** |
| 139 | CHSH | 0.2695 | **** | 0.5330 | -17.4570 | **** | **** | **** | **** | **** | **** |
| 140 | CSH | **** | **** | -0.1180 | -16.5064 | **** | **** | **** | **** | **** | **** |
| 141 | aC-SH | **** | **** | 0.7935 | -16.4699 | **** | 3.7519 | **** | **** | **** | **** |
| 142 | -SH except as above | 0.5489 | **** | -0.4044 | -10.7037 | **** | 2.6753 | **** | **** | **** | **** |
| 143 | $CH_3S$ | 0.2836 | 2.1003 | 0.4241 | -16.2229 | 0.9628 | 2.6895 | **** | **** | **** | **** |
| 144 | $CH_2S$ | 0.5397 | -1.4511 | 0.3870 | -16.2453 | -0.5093 | **** | **** | **** | **** | **** |
| 145 | CHS | 0.7313 | 9.4995 | 0.5624 | -16.1516 | -0.7805 | -1.2947 | **** | **** | **** | **** |
| 146 | CS | 1.1701 | **** | 0.3257 | -16.5664 | **** | **** | **** | **** | **** | **** |
| 147 | aC-S- | 1.0083 | -0.5150 | 0.4494 | -15.7488 | -1.0166 | **** | **** | **** | **** | **** |
| 148 | SO | -1.5826 | -2.4432 | 0.1023 | -14.7151 | -1.5532 | -0.2469 | -0.5536 | **** | **** | **** |
| 149 | $SO_2$ | 2.5425 | **** | 0.3801 | -21.7306 | -1.0301 | 2.3321 | **** | **** | **** | **** |
| 150 | $SO_3$ (sulfite) | **** | **** | -0.1636 | -31.4604 | **** | **** | **** | **** | **** | **** |
| 151 | $SO_3$ (Sulfonate) | 0.4513 | **** | 0.3305 | -27.2267 | **** | **** | **** | **** | **** | **** |
| 152 | $SO_4$ (Sulfate) | **** | **** | 0.8082 | -32.5350 | **** | 2.9901 | **** | **** | **** | **** |
| 153 | aC-SO | -0.5034 | **** | 0.7318 | -21.6327 | -2.4918 | **** | **** | **** | **** | **** |
| 154 | aC-$SO_2$ | 0.4065 | -0.8851 | 0.0966 | -26.3450 | -2.0960 | **** | **** | **** | **** | **** |
| 155 | PH (phosphine) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 156 | P (Phospine) | **** | **** | 0.6540 | **** | **** | **** | **** | **** | **** | **** |
| 157 | $PO_3$ (Phospite) | **** | **** | 0.5089 | **** | -1.6943 | **** | **** | **** | **** | **** |
| 158 | $PHO_3$ (Phosponate) | 0.3773 | **** | 0.0938 | -25.8155 | **** | **** | **** | **** | **** | **** |
| 159 | $PO_3$ (Phosponate) | 0.0507 | 2.3731 | 1.3895 | -26.0493 | -2.9814 | **** | **** | **** | **** | **** |
| 160 | $PHO_4$ (Phospate) | 1.5332 | **** | -0.0158 | -30.8958 | -1.7167 | **** | **** | **** | **** | **** |
| 161 | $PO_4$ (Phospate) | 0.0473 | 2.3370 | 1.8277 | -30.9346 | -2.0276 | **** | **** | **** | **** | **** |
| 162 | aC-$PO_4$ | 2.0153 | 4.5391 | 1.1477 | -36.6047 | -1.2336 | 0.1450 | **** | **** | **** | **** |
| 163 | aC-P | **** | **** | 0.6623 | **** | **** | **** | **** | **** | **** | **** |
| 164 | $CO_3$ (Carbonate) | **** | **** | 0.4013 | -20.0014 | **** | **** | 1.5293 | **** | **** | **** |
| 165 | $C_2H_3O$ | 0.8118 | **** | 0.3448 | -14.1859 | **** | 2.0487 | 0.7572 | **** | **** | **** |
| 166 | $C_2H_2O$ | -1.4431 | **** | 0.0268 | -14.0802 | **** | **** | **** | **** | **** | **** |
| 167 | $C_2HO$ | **** | **** | 0.2419 | -15.2773 | **** | **** | **** | **** | **** | **** |
| 168 | $CH_2$ (cyclic) | 0.1038 | -0.2352 | 0.0255 | -4.8771 | 0.2427 | 0.2430 | 0.0436 | **** | **** | **** |
| 169 | CH (cyclic) | 0.3421 | 1.1919 | 0.3515 | -5.0340 | -0.2328 | -0.1306 | 0.0555 | **** | **** | **** |
| 170 | C (cyclic) | -0.6996 | 0.6314 | 0.3590 | -4.5355 | -0.0624 | -2.1325 | 0.0103 | **** | **** | **** |
| 171 | CH=CH (cyclic) | 0.2634 | 0.5840 | 0.0512 | -9.1129 | 0.6403 | 0.9258 | -0.1761 | **** | **** | **** |
| 172 | CH=C (cyclic) | 0.9929 | -2.9859 | 0.2542 | -9.0957 | -1.0636 | -1.0568 | -0.6613 | **** | **** | **** |
| 173 | C=C (cyclic) | 1.4317 | 0.2367 | 0.3198 | -9.3336 | -0.2606 | -1.2451 | **** | **** | **** | **** |
| 174 | $CH_2$=C (cyclic) | 1.1079 | 2.1640 | 0.6046 | -8.6263 | 0.1438 | -2.4611 | -0.6120 | **** | **** | **** |
| 175 | NH (cyclic) | -0.5856 | 0.3105 | 0.1081 | -4.9664 | -0.4746 | 3.6453 | **** | **** | **** | **** |
| 176 | N (cyclic) | -0.5091 | -3.7536 | 0.1248 | -4.4756 | 0.1110 | -0.1597 | -0.1781 | **** | **** | **** |
| 177 | CH=N (cyclic) | 0.0220 | 0.7125 | -0.0103 | -9.3002 | 0.3716 | **** | **** | **** | **** | **** |
| 178 | C=N (cyclic) | 0.1731 | 2.6967 | 0.1735 | -9.3647 | -0.0603 | **** | **** | **** | **** | **** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 179 | O (cyclic) | -0.3445 | -0.3108 | 0.0003 | -5.0128 | -0.7104 | 1.2330 | 0.1823 | **** | **** | **** |
| 180 | CO (cyclic) | -0.2591 | -0.8055 | 0.0617 | -9.7579 | -0.1690 | 1.9059 | 0.3317 | **** | **** | **** |
| 181 | S (cyclic) | 1.3798 | -1.2604 | 0.0539 | -11.0518 | 0.1784 | -3.3757 | **** | **** | **** | **** |
| 182 | $SO_2$ (cyclic) | **** | 0.0804 | 0.2151 | -21.4596 | -0.5954 | **** | **** | **** | **** | **** |
| 183 | >NH | 0.0556 | -0.4808 | 0.3294 | -4.6200 | -0.5666 | **** | **** | **** | **** | **** |
| 184 | -O- | -0.8862 | 1.7786 | 0.5619 | -5.1116 | -0.5432 | -6.3984 | **** | **** | **** | **** |
| 185 | -S- | 1.0294 | 2.4264 | 0.1568 | -11.7368 | -0.3619 | **** | **** | **** | **** | **** |
| 186 | >CO | 1.3751 | 0.8472 | 0.0875 | -9.6930 | 0.4503 | 0.5076 | 0.1421 | **** | **** | **** |
| 187 | $PO_2$ | **** | **** | 0.1658 | **** | **** | **** | **** | **** | **** | **** |
| 188 | CH-N | **** | 2.3606 | -0.1676 | -9.1669 | **** | **** | **** | **** | **** | **** |
| 189 | SiHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 190 | SiO | **** | **** | -0.0142 | -16.1963 | -0.3192 | -2.9881 | **** | **** | **** | **** |
| 191 | $SiH_2$ | **** | **** | -0.2242 | **** | **** | **** | **** | **** | **** | **** |
| 192 | SiH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 193 | Si | **** | **** | 0.3379 | -11.4006 | -1.4109 | **** | **** | **** | **** | **** |
| 194 | $(CH_3)_3N$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 195 | N=N | 0.7152 | **** | 0.0892 | -9.4458 | **** | **** | **** | **** | **** | **** |
| 196 | $C_{cyclic}$=N- | 1.1742 | **** | 0.5802 | -9.4880 | -0.6167 | **** | **** | **** | **** | **** |
| 197 | $C_{cyclic}$=CH- | 1.9779 | 1.0355 | 0.6675 | -8.9669 | -0.6699 | 0.9004 | **** | **** | **** | **** |
| 198 | $C_{cyclic}$=NH | **** | **** | -0.0374 | -9.9059 | **** | **** | **** | **** | **** | **** |
| 199 | N=O | -0.0312 | 0.7543 | 0.3193 | -9.8426 | -0.2232 | **** | **** | **** | **** | -0.0775 |
| 200 | $C_{cyclic}$=C | 1.2298 | **** | 0.4824 | -9.0043 | -1.2043 | **** | 0.0712 | **** | **** | **** |
| 201 | P=O | 1.1226 | -0.5256 | 0.8567 | -16.1074 | -1.4127 | **** | **** | **** | **** | **** |
| 202 | N=N | **** | 1.1355 | 0.3544 | -10.8546 | -0.2014 | **** | **** | **** | **** | **** |
| 203 | C=NH | -0.7564 | **** | -0.3121 | -10.0237 | -0.6413 | **** | **** | **** | **** | **** |
| 204 | >C=S | 0.8375 | -0.3582 | 0.2675 | -15.7517 | 0.0854 | **** | **** | **** | **** | **** |
| 205 | aC-CON | 0.3392 | -2.9838 | 0.2853 | -17.8813 | -2.7177 | **** | **** | **** | **** | **** |
| 206 | aC=O | -0.0010 | **** | -0.1490 | -10.0219 | -0.4153 | **** | **** | **** | **** | **** |
| 207 | aN- | -0.3370 | 0.6885 | 0.2069 | -4.2632 | -1.1185 | **** | **** | **** | **** | **** |
| 208 | -Na | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 209 | -K | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 210 | HCONH | **** | **** | -0.1282 | -14.4140 | -1.0365 | **** | **** | **** | **** | **** |
| 211 | CHOCH | 1.3384 | 0.5499 | 0.8732 | -14.0900 | **** | **** | **** | **** | **** | **** |
| 212 | $C_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 213 | $SiH_3$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 214 | $SiH_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 215 | CH=C=CH | **** | **** | **** | -13.3858 | **** | **** | **** | **** | **** | **** |
| 216 | CH=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 217 | OP(=S)O | 1.0213 | 3.8226 | 1.4105 | -33.4664 | 0.0143 | 0.7109 | **** | **** | **** | **** |
| 218 | R | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 219 | $CF_{2cyclic}$ | **** | **** | 0.3368 | -17.7706 | **** | **** | **** | 1.0010 | **** | **** |
| 220 | $CF_{cyclic}$ | **** | **** | -0.1246 | -11.4927 | **** | **** | **** | **** | **** | **** |

[a] The symbols $LC50(FM)_{1i}$, $LC50(DM)_{1i}$, $LD50_{1i}$, $LogWs_{1i}$, $BCF_{1i,}$ $PEL_{1i,}$ $PCO_{1i}$, $GWP_{1i}$, $ODP_{1i}$,and $AP_{1i}$ represent the contributions ($C_i$) of the first-order groups for the corresponding properties.

**Table S5 (continued). MG Method Based Property Models Analysed Using Simultaneous Regression Method: First-Order Groups and their Contributions [a] for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| | Group | $EUA_{C\,1i}$ | $EUA_{NC\,1i}$ | $ERA_{C\,1i}$ | $ERA_{NC\,1i}$ | $EFW_{C\,1i}$ | $EFW_{NC\,1i}$ | $ESW_{C\,1i}$ | $ESW_{NC\,1i}$ | $ENS_{C\,1i}$ | $ENS_{NC\,1i}$ | $EAS_{C\,1i}$ | $EAS_{NC\,1i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $CH_3$ | 0.2768 | -0.3889 | 0.4186 | -0.8784 | 0.7393 | -0.3816 | -0.3653 | -1.0175 | 0.6318 | 0.1850 | 0.6353 | 0.2561 |
| 2 | $CH_2$ | -0.1334 | -0.0407 | -0.1321 | -0.0319 | -0.1459 | -0.0700 | -0.1772 | -0.0575 | 0.0849 | 0.1648 | 0.0411 | 0.1062 |
| 3 | CH | 0.0802 | 0.0707 | 0.3152 | 0.5123 | 0.1677 | 0.2100 | 1.0114 | 0.3785 | 0.3251 | -0.2407 | -0.0273 | -0.5895 |
| 4 | C | 0.9788 | 0.9223 | 0.2656 | 1.7157 | -0.3090 | -0.0359 | 1.9303 | 0.3055 | 0.2978 | -0.9176 | -0.6505 | -1.1619 |
| 5 | $CH_2$=CH | -0.3402 | -1.3987 | 0.8962 | -2.0681 | 0.9410 | -1.6118 | -0.2764 | -1.1415 | 0.9785 | -0.5469 | 1.1985 | 0.0661 |
| 6 | CH=CH | -1.3145 | -2.1775 | 1.0385 | -0.5744 | 0.4864 | -1.2323 | 0.2899 | -0.6143 | 0.5527 | -1.8356 | 0.6069 | -0.7348 |
| 7 | $CH_2$=C | -1.3393 | -1.3567 | 1.6881 | -0.3947 | 1.1356 | -1.0631 | 0.7675 | 0.2210 | 1.9580 | -1.1219 | 2.0215 | -0.0034 |
| 8 | CH=C | -3.0812 | -1.8284 | 0.2217 | -0.8211 | -0.4232 | -1.1149 | 0.2755 | 0.2933 | 0.2432 | -3.2576 | 0.1952 | -1.7401 |
| 9 | C=C | -3.0673 | -2.3768 | 1.4520 | -0.3112 | -1.5124 | -1.1376 | 0.8643 | 1.5679 | -1.1132 | -4.2931 | -1.3559 | -2.0206 |
| 10 | $CH_2$=C=CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | $CH_2$=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 12 | C=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | CH≡C | -4.0744 | -1.7053 | -0.4115 | -2.0220 | -0.1562 | -1.8383 | -0.5146 | -1.4962 | 0.1162 | -0.6296 | -0.2172 | -0.6574 |
| 14 | C≡C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 15 | aCH | 0.1318 | -0.0824 | 0.1728 | -0.3183 | 0.2732 | -0.1185 | -0.0798 | -0.2644 | 0.2984 | 0.1383 | 0.3088 | 0.1538 |
| 16 | aC fused with aromatic ring | -0.1343 | 0.3895 | -0.5355 | 0.5155 | -0.5066 | 0.5060 | -0.4842 | 0.3773 | -0.1470 | 0.5247 | -0.3689 | 2.2520 |
| 17 | aC fused with non-aromatic ring | -0.7234 | 0.1826 | -1.0425 | 0.6396 | -0.9302 | 0.7453 | -1.0106 | 0.5475 | -0.8140 | 0.5110 | -0.9460 | 1.9490 |
| 18 | aC except as above | -0.8413 | -0.0646 | -0.7811 | 0.6090 | -1.2107 | -0.2465 | -1.0751 | -0.1323 | -0.5215 | 0.0358 | -0.5195 | -0.4922 |
| 19 | aN in aromatic ring | 0.0545 | -0.5575 | -0.1527 | -0.8743 | 0.2922 | -0.4505 | 0.0351 | -0.5408 | 0.1579 | -0.3233 | 0.0617 | -0.2065 |
| 20 | aC-$CH_3$ | -0.5357 | -0.0847 | 0.1353 | 0.0871 | -0.1590 | 0.1514 | -0.5494 | 0.0327 | -0.0284 | 0.6381 | 0.0242 | 0.7085 |
| 21 | aC-$CH_2$ | -0.5009 | 0.9054 | -0.9434 | 1.1561 | -1.0858 | 0.4515 | -0.6868 | 0.8390 | -0.9057 | 0.2123 | -0.7848 | 0.1967 |
| 22 | aC-CH | -0.6200 | 0.0515 | -1.2974 | 0.9275 | -2.2675 | -0.5408 | -0.3939 | -0.1009 | -1.4835 | -0.8336 | -1.4515 | -1.1191 |
| 23 | aC-C | -3.3208 | 0.2718 | -3.9393 | 1.5295 | -6.1374 | -0.5875 | -3.4758 | -0.0850 | -5.2596 | -1.3713 | -5.3546 | -1.8340 |
| 24 | aC-CH=$CH_2$ | 0.1127 | 0.0416 | 0.2302 | 0.5380 | 1.4500 | 1.4195 | -0.1057 | 0.7205 | 0.6742 | 1.6855 | 1.0702 | 1.8907 |
| 25 | aC-CH=CH | 1.0394 | 0.2028 | 1.9524 | -1.0455 | 1.2025 | -0.2276 | 1.4654 | -0.1484 | 1.6964 | -0.7658 | 1.5229 | -1.7432 |
| 26 | aC-C=$CH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 27 | aC-C≡CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 28 | aC-C≡C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 29 | OH | 0.4194 | 0.5362 | 0.2433 | -0.8277 | 0.8887 | 0.3100 | 0.2217 | 0.5225 | 0.3945 | 0.3581 | 0.6619 | 0.6345 |
| 30 | aC-OH | 0.5318 | -0.1766 | 0.4224 | -0.6903 | 0.5713 | -0.4110 | 0.7130 | 0.3440 | 0.9767 | 0.4142 | 1.0735 | 0.3160 |
| 31 | COOH | 0.9436 | -0.3737 | 0.9872 | -1.1003 | 1.5695 | -0.0513 | 1.0489 | 0.7155 | 1.0456 | 0.8499 | 0.9371 | 0.2633 |
| 32 | aC-COOH | -1.9180 | -0.1662 | 0.6256 | -0.9179 | 1.1534 | 0.1616 | 1.2207 | 1.9706 | 1.3280 | 0.2411 | 1.2017 | 0.1223 |
| 33 | $CH_3CO$ | 0.0905 | 1.3373 | -0.6548 | -0.1061 | **** | 1.2797 | -0.5174 | 0.8365 | -0.0497 | 2.0562 | **** | 2.1799 |
| 34 | $CH_2CO$ | -3.7621 | **** | -3.7891 | **** | -3.6144 | **** | 0.8647 | **** | -4.7374 | **** | -4.5274 | **** |
| 35 | CHCO | **** | -0.1163 | **** | 0.1359 | **** | -0.3423 | **** | 0.2564 | **** | -0.4092 | **** | -1.0036 |
| 36 | CCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 37 | aC-CO | -3.6693 | 0.1230 | -3.2999 | -0.2251 | -2.1671 | -0.3041 | -2.1170 | 0.0400 | -3.3091 | -0.0111 | -3.0206 | -0.4405 |
| 38 | CHO | 0.7415 | -1.0946 | 0.3433 | -1.4468 | 0.7471 | -0.8811 | -0.3255 | -0.4580 | 0.4532 | 0.2603 | 0.5964 | 0.1867 |
| 39 | aC-CHO | 1.5855 | -0.1253 | 1.1053 | -0.3669 | 1.1466 | -0.2715 | 1.0434 | -0.1487 | 0.8989 | 0.6655 | 0.9465 | 0.3748 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | $CH_3COO$ | 0.4512 | 0.9230 | 0.4035 | 0.0309 | 0.6553 | 1.1335 | 0.1280 | 0.2915 | 0.4063 | 1.7726 | 0.5420 | 1.4246 |
| 41 | $CH_2COO$ | 1.1595 | 1.2031 | 0.1357 | 0.7204 | 0.4287 | 1.4075 | 1.5236 | 2.3643 | -0.9216 | 1.2035 | -0.0509 | 1.0455 |
| 42 | CHCOO | **** | 0.3293 | **** | 1.0144 | **** | 2.4473 | **** | 2.5891 | **** | 0.6452 | **** | 0.1441 |
| 43 | CCOO | 0.1221 | **** | -0.2906 | **** | -0.8719 | **** | 0.3265 | **** | -0.5178 | **** | -0.8982 | **** |
| 44 | HCOO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 45 | aC-COO | -0.4819 | 0.4459 | -0.4124 | 0.4075 | -0.4451 | 0.5003 | -0.1062 | 1.3013 | -0.2463 | 0.4928 | -0.4509 | -0.0204 |
| 46 | aC-OOCH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 47 | aC-OOC | **** | 0.5722 | 2.5567 | 0.2878 | -0.7154 | 0.5013 | 6.7362 | 1.0671 | 1.0188 | 0.2805 | 1.1582 | -0.0771 |
| 48 | COO except as above | 0.2977 | -1.5617 | -0.3523 | -1.4325 | -0.1730 | -1.8607 | -0.1199 | -0.8076 | -0.1119 | -0.3508 | -0.0340 | -0.0515 |
| 49 | $CH_3O$ | 0.6256 | 0.0811 | -0.2026 | -0.9191 | 0.6929 | -0.0290 | -0.7151 | -0.2724 | 0.3522 | 0.3233 | 0.5985 | 0.5144 |
| 50 | $CH_2O$ | 0.0748 | 0.1975 | 0.4215 | -0.2890 | 0.1084 | 0.1516 | -1.0241 | 0.1937 | 0.2143 | -0.6273 | 0.1613 | -0.3999 |
| 51 | CH-O | **** | -0.0821 | **** | -0.4530 | **** | -0.6497 | **** | 0.5274 | **** | -1.2816 | **** | -1.7349 |
| 52 | C-O | -5.2568 | **** | 0.9916 | **** | -0.5381 | **** | -3.1651 | **** | -1.4749 | **** | -2.4229 | **** |
| 53 | aC-O | -0.0802 | -0.2859 | -0.0713 | -0.0853 | -0.4911 | -0.2147 | 0.2796 | 0.6630 | -0.3339 | -0.9133 | -0.2773 | -0.8960 |
| 54 | $CH_2NH_2$ | 2.6349 | **** | 2.1445 | **** | 2.7445 | **** | 2.2313 | **** | 2.2150 | **** | 2.5740 | **** |
| 55 | $CHNH_2$ | -0.6207 | **** | -0.4900 | **** | -0.4117 | **** | -0.1089 | **** | -0.9221 | **** | -0.6954 | **** |
| 56 | $CNH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 57 | $CH_3NH$ | 0.6012 | 3.5226 | 1.1697 | 0.8357 | 0.5009 | 3.0770 | 0.0504 | 3.3927 | 0.5130 | 4.1468 | 0.6086 | 2.6084 |
| 58 | $CH_2NH$ | -0.0418 | 1.1184 | 0.2209 | 1.8720 | -0.2442 | **** | 0.2238 | -0.4997 | 0.0384 | **** | 0.0274 | **** |
| 59 | CHNH | **** | -1.6773 | **** | -1.5280 | **** | -1.6357 | **** | -1.2850 | **** | -3.0506 | **** | -3.2766 |
| 60 | $CH_3N$ | -0.6436 | 0.9099 | -0.8553 | -0.1924 | -0.8106 | 0.3857 | -0.6728 | 1.2439 | -0.6269 | -0.2579 | -0.5892 | 0.0841 |
| 61 | $CH_2N$ | 0.4475 | -2.0098 | 0.0993 | -2.4756 | -0.3946 | -2.7768 | 0.6252 | -2.9607 | -0.2504 | -2.2740 | -0.3272 | -5.6684 |
| 62 | $aC\text{-}NH_2$ | 0.4078 | -0.0630 | 0.3147 | -0.3145 | 0.2777 | -0.2788 | 0.1703 | -0.0925 | 0.3313 | -0.0620 | 0.5484 | 0.0665 |
| 63 | aC-NH | -0.4107 | 0.4817 | -0.7748 | 0.3599 | -1.1879 | 0.0024 | -0.3339 | 0.7840 | -0.8853 | -0.2743 | -0.8825 | -0.4021 |
| 64 | aC-N | -0.4531 | 1.8888 | -0.3214 | 2.3247 | -1.6764 | 1.1305 | -0.0422 | 1.7549 | -1.2858 | -0.0682 | -1.5082 | -0.5018 |
| 65 | $NH_2$ except as above | -0.3084 | -0.6762 | -0.5831 | -1.2756 | -0.0303 | -1.2682 | -0.4261 | 0.0098 | -0.1112 | 0.1261 | -0.0458 | 0.3308 |
| 66 | CH=N | -0.9032 | -0.8004 | 0.3303 | -0.9608 | -0.0639 | -0.9532 | -0.5617 | -1.0261 | 0.2969 | -0.0600 | 0.3659 | -0.4586 |
| 67 | C=N | 0.4874 | 0.7296 | 0.0018 | 1.3884 | -0.1105 | 0.1401 | 0.2676 | 0.2538 | 0.8550 | 0.9988 | 0.7018 | 0.1943 |
| 68 | $CH_2CN$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 69 | CHCN | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 70 | CCN | -0.6082 | **** | -0.8192 | **** | -1.0567 | **** | 0.2996 | **** | -1.4462 | **** | -0.9915 | **** |
| 71 | aC-CN | 1.0739 | -0.1684 | 0.9127 | -0.4935 | 1.2596 | -0.0803 | 1.2767 | -0.0518 | 1.2816 | 0.0372 | 1.2918 | 0.2045 |
| 72 | CN except as above | 0.8164 | -0.2555 | -0.6385 | -1.4802 | 0.7394 | -0.2068 | 0.3631 | -0.5179 | 0.9006 | 0.1745 | 0.9294 | 0.7612 |
| 73 | $CH_2NCO$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 74 | CHNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 75 | CNCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 76 | aC-NCO | 1.1330 | -2.1259 | 0.4514 | -2.3525 | 0.5267 | -2.0876 | -0.4091 | -2.3986 | 1.1762 | -1.2588 | 1.1214 | -1.4601 |
| 77 | $CH2NO_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 78 | $CHNO_2$ | -0.7932 | -1.2349 | -2.0288 | -2.3937 | -1.2622 | -1.0859 | -1.9514 | -1.5630 | -1.6619 | -1.1247 | -1.2586 | -1.1625 |
| 79 | $CNO_2$ | -0.9724 | **** | 1.4716 | **** | -0.2284 | **** | -2.2112 | **** | 1.3183 | **** | 0.7714 | **** |
| 80 | $aC\text{-}NO_2$ | 0.1245 | -0.7880 | -0.5036 | -1.1595 | -0.0264 | -0.6468 | -0.4483 | -0.4066 | 0.1379 | -0.1310 | 0.3479 | -0.1339 |
| 81 | $NO_2$ except as above | -0.0683 | 0.8138 | -1.1988 | -0.2626 | -0.1981 | 0.8788 | -0.5692 | 0.2233 | -0.7637 | 2.1380 | -0.4482 | 2.0678 |
| 82 | ONO | 0.0475 | **** | -0.7753 | **** | 0.0152 | **** | -1.8329 | **** | -0.3701 | **** | 0.0539 | **** |
| 83 | $ONO_2$ | 0.4673 | **** | -0.0272 | **** | 0.4472 | **** | -0.1621 | **** | 0.2459 | **** | 0.4794 | **** |
| 84 | $HCON(CH_2)_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 85 | $HCONHCH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | $CONH_2$ | 0.6701 | -0.6830 | 0.4481 | -0.6628 | 1.4781 | -0.7170 | 1.9978 | -0.3006 | 1.2097 | 0.1667 | 1.3995 | 0.6316 |
| 87 | $CONHCH_3$ | -0.3275 | -1.0584 | -1.4288 | -2.0736 | -0.1886 | -1.5198 | -0.6454 | -1.2248 | -1.0779 | -0.3723 | -0.8224 | -0.7812 |
| 88 | $CONHCH_2$ | 1.9906 | 1.8243 | 0.9711 | 0.8861 | 0.7879 | 2.0946 | 2.5279 | 1.7462 | 1.6381 | 2.5388 | 1.6578 | 2.4492 |
| 89 | $CON(CH_3)_2$ | 1.3855 | -0.1214 | -0.5104 | -1.2591 | -0.2235 | -0.7867 | 0.1119 | -0.5949 | -0.9044 | 0.2034 | -0.1317 | 0.1891 |
| 90 | $CONCH_3CH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 91 | $CON(CH_2)_2$ | -0.9625 | 0.2621 | -1.5819 | 0.2178 | -2.1596 | -0.3609 | 0.1099 | 0.4682 | -1.5442 | -0.7156 | -1.3442 | -1.2618 |
| 92 | CONHCO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 93 | CONCO | **** | -0.3656 | **** | -0.0544 | **** | -0.2574 | **** | 1.3425 | **** | -0.6528 | **** | -0.6126 |
| 94 | aC-$CONH_2$ | 1.3190 | **** | 0.4221 | **** | 1.2138 | **** | 0.5586 | **** | 0.6842 | **** | 1.0114 | **** |
| 95 | aC-NH(CO)H | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 96 | aC-N(CO)H | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 97 | aC-CONH | -3.6015 | 1.3725 | -2.4897 | 1.0603 | -1.3548 | 1.2428 | -0.8792 | 1.3704 | -2.5327 | 0.7981 | -1.7480 | 0.7555 |
| 98 | aC-NHCO | 0.3259 | 0.3311 | 0.0376 | 0.8914 | -0.2426 | 0.7337 | 0.9831 | 1.1069 | -0.3504 | 0.6114 | -0.0668 | 0.5521 |
| 99 | aC-(N)CO | -0.8438 | 0.8189 | -0.9527 | 1.2103 | -2.1512 | 0.6037 | 0.1744 | 2.7438 | -0.9767 | -0.2882 | -0.9029 | -0.4164 |
| 100 | NHCONH | 0.9715 | **** | 0.8013 | **** | 1.0687 | **** | 1.6118 | **** | 0.6872 | **** | 0.7296 | **** |
| 101 | $NH_2CONH$ | 1.1915 | **** | -0.0835 | **** | 0.8273 | **** | 0.9456 | **** | 0.5512 | **** | 0.6840 | **** |
| 102 | $NH_2CON$ | 0.3622 | **** | -1.0023 | **** | -0.3280 | **** | 0.0298 | **** | -0.2842 | **** | -0.3256 | **** |
| 103 | NHCON | 1.9570 | 0.5198 | 0.1885 | 1.0282 | 0.7393 | 1.0205 | 1.8051 | 3.8658 | 0.8960 | -1.7350 | 0.5601 | -1.4133 |
| 104 | NCON | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 105 | aC-$NHCONH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 106 | aC-NHCONH | **** | 0.8147 | 1.6760 | 1.1664 | 0.6501 | 0.4940 | **** | 0.3930 | 1.4306 | -0.1933 | 1.6286 | -0.1363 |
| 107 | NHCO except as above | -0.1165 | -0.6416 | -0.4101 | -0.1778 | -0.1483 | -0.5051 | 0.7002 | -0.3826 | -0.1430 | -0.7959 | -0.0158 | -0.3768 |
| 108 | $CH_2Cl$ | -0.3133 | -0.9231 | -0.3902 | -1.5409 | -0.0012 | -0.8957 | -0.7134 | -1.7334 | -0.0092 | -0.3747 | 0.0485 | -0.0864 |
| 109 | CHCl | 0.7229 | 0.2038 | -0.3138 | -0.2775 | 0.4587 | -0.4287 | -1.0845 | -0.5868 | 0.1700 | -0.1056 | 0.3100 | 0.1483 |
| 110 | CCl | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 111 | $CHCl_2$ | 0.4190 | -0.9871 | -0.1853 | -2.4110 | 0.6353 | -1.0667 | -1.0422 | -1.6383 | 0.3711 | -0.3263 | 0.4982 | -0.2269 |
| 112 | $CCl_2$ | **** | -0.4318 | **** | -0.5196 | **** | -0.6916 | **** | -0.5364 | **** | -0.4357 | **** | -0.2253 |
| 113 | $CCl_3$ | 0.3605 | -1.2506 | -0.4669 | -2.4145 | 0.2786 | -1.5568 | -1.8568 | -2.2684 | -0.1843 | -0.6057 | 0.0910 | -0.5476 |
| 114 | $CH_2F$ | 1.7119 | 0.3666 | 1.0533 | -0.8101 | 1.8094 | 0.7354 | 0.6295 | 0.0059 | 1.4211 | 0.4477 | 1.6553 | 0.6039 |
| 115 | CHF | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 116 | CF | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 117 | $CHF_2$ | **** | 0.0703 | **** | -0.4804 | **** | 0.1719 | **** | -1.1833 | **** | 1.9698 | **** | 1.6025 |
| 118 | $CF_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 119 | $CF_3$ | 0.9473 | 0.8895 | 0.1773 | -0.3682 | 1.0720 | 0.3204 | -1.3467 | -1.0247 | 0.5889 | 1.5174 | 0.8026 | 1.7167 |
| 120 | $CCl_2F$ | **** | -1.4099 | **** | -2.1593 | **** | -1.7969 | **** | -3.1682 | **** | -1.4242 | **** | -1.1754 |
| 121 | HCClF | **** | 1.7393 | **** | 0.7917 | **** | 1.3080 | **** | -0.1021 | **** | 1.6004 | **** | 1.9043 |
| 122 | $CClF_2$ | **** | 0.7293 | **** | -0.9851 | **** | 0.8903 | **** | 0.0847 | **** | 1.4088 | **** | 1.5158 |
| 123 | aC-Cl | -0.0445 | -0.2890 | -0.3069 | -0.7079 | -0.1654 | -0.4874 | -0.8314 | -0.8213 | -0.1399 | -0.1050 | -0.0633 | -0.0999 |
| 124 | aC-F | -0.9985 | -0.2195 | -1.2055 | -0.4714 | -0.7630 | -0.2717 | -1.4596 | 0.0464 | -0.6826 | 0.0476 | -1.2152 | 0.2171 |
| 125 | aC-I | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 126 | aC-Br | 0.2093 | -0.2519 | 0.0753 | -0.6226 | 0.3176 | -0.2506 | -0.0476 | -0.4843 | 0.4777 | -0.0766 | 0.2864 | -0.1032 |
| 127 | -I except as above | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 128 | -Br except as above | -0.1961 | -0.3216 | -0.1708 | -1.2917 | 0.0232 | -0.3045 | -0.9789 | -1.1443 | -0.2296 | 0.0657 | -0.0047 | 0.3657 |
| 129 | -F except as above | 0.6207 | -0.7864 | -0.5794 | -2.2689 | 0.6589 | -0.5031 | -1.9094 | -0.9024 | -0.1512 | 0.0691 | 0.2341 | 0.1207 |
| 130 | -Cl except as above | -1.9393 | -0.1480 | -0.7165 | -2.0875 | -0.3256 | -0.0976 | -1.4034 | -1.2269 | 0.0434 | 0.3658 | -0.4433 | 0.5279 |
| 131 | CHNOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 132 | CNOH | 0.1658 | **** | -1.0768 | **** | -0.6315 | **** | 0.2332 | **** | -0.3801 | **** | -0.0535 | **** |
| 133 | aC-CHNOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 134 | $OCH_2CH_2OH$ | 2.3891 | -0.0244 | 1.7788 | -1.2602 | 2.1537 | -0.4130 | 2.4413 | 0.9540 | 1.6546 | 0.2117 | 1.8160 | 0.3751 |
| 135 | $OCHCH_2OH$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 136 | $OCH_2CHOH$ | **** | 2.4833 | **** | 0.8375 | **** | 2.0173 | **** | 4.0842 | **** | 1.7767 | **** | 1.9349 |
| 137 | -O-OH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 138 | $CH_2SH$ | -0.6250 | **** | -1.9924 | **** | -1.1190 | **** | -0.9189 | **** | -0.0570 | **** | 0.1574 | **** |
| 139 | CHSH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 140 | CSH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 141 | aC-SH | -1.2297 | **** | -3.7114 | **** | -2.1847 | **** | -0.4976 | **** | -1.8963 | **** | -1.8779 | **** |
| 142 | -SH except as above | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 143 | $CH_3S$ | 6.3958 | -0.1547 | 0.9366 | -0.3871 | 1.3187 | 1.0294 | 1.6307 | -0.3173 | 1.1446 | -0.4689 | 1.1055 | -0.2242 |
| 144 | $CH_2S$ | -0.3057 | -0.0492 | -1.0545 | -0.3098 | -1.3645 | -0.2406 | -0.6499 | -0.3212 | -1.0185 | -0.2384 | -1.0862 | -0.3142 |
| 145 | CHS | **** | -1.3648 | **** | -1.1438 | **** | -2.1457 | **** | -3.5270 | **** | -3.4275 | **** | -3.2466 |
| 146 | CS | 7.6552 | -0.9007 | 3.4256 | 3.9642 | 2.6418 | -1.3869 | 3.8903 | -0.5260 | 1.0593 | -2.3698 | 2.8471 | -2.3476 |
| 147 | aC-S- | **** | 0.2217 | -1.0230 | 0.1749 | **** | -0.0818 | -1.3298 | 0.1865 | **** | -0.2047 | -1.0183 | -0.2664 |
| 148 | SO | 1.7021 | 4.3538 | 0.9399 | 2.4815 | 0.3344 | 4.3129 | 0.3485 | 4.3583 | -0.3864 | 3.2094 | -0.0328 | 0.2414 |
| 149 | $SO_2$ | **** | -0.3824 | **** | -0.0907 | **** | 0.0637 | **** | 0.4310 | **** | 1.5538 | **** | 1.1670 |
| 150 | $SO_3$ (sulfite) | -0.0827 | **** | -0.3225 | -1.3680 | -1.0811 | -0.6689 | -1.5796 | 0.8753 | -0.7085 | **** | -0.4129 | 0.3305 |
| 151 | $SO_3$ (Sulfonate) | -0.2533 | **** | -1.7071 | **** | -0.9481 | **** | 0.5266 | **** | -1.0911 | **** | -0.7997 | **** |
| 152 | $SO_4$ (Sulfate) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 153 | aC-SO | **** | -0.6036 | **** | -0.7755 | **** | -0.7480 | **** | 0.6670 | **** | -1.0006 | **** | -0.8386 |
| 154 | aC-$SO_2$ | -2.3992 | -0.2423 | -2.8465 | -0.3578 | -2.9877 | -1.4779 | -1.6383 | -1.5973 | -3.2071 | -0.1751 | -3.0574 | -0.8298 |
| 155 | PH (phosphine) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 156 | P (Phospine) | **** | -1.0674 | **** | -0.5893 | **** | -2.1490 | **** | -1.7547 | **** | -0.9683 | **** | -0.7173 |
| 157 | $PO_3$ (Phospite) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 158 | $PHO_3$ (Phosponate) | 0.8052 | **** | -0.3669 | **** | -0.0439 | **** | 1.2624 | **** | -0.1360 | **** | 0.1555 | **** |
| 159 | $PO_3$ (Phosponate) | 1.3267 | 0.0298 | 0.7963 | -0.0321 | -0.4619 | 0.1061 | 2.3937 | 2.0215 | 0.2316 | -0.8312 | 0.2283 | -0.2818 |
| 160 | $PHO_4$ (Phospate) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 161 | $PO_4$ (Phospate) | 1.6283 | -0.5208 | -0.2120 | 0.0581 | 0.3359 | -0.6785 | 2.0237 | 1.0577 | 0.5070 | -0.5538 | 0.2061 | -1.0503 |
| 162 | aC-$PO_4$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 163 | aC-P | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 164 | $CO_3$ (Carbonate) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 165 | $C_2H_3O$ | -0.3133 | -5.1385 | -1.2673 | -4.7155 | 1.1448 | -4.7775 | -0.1153 | -7.7255 | 0.0260 | -3.6274 | 0.9980 | -4.0098 |
| 166 | $C_2H_2O$ | **** | -1.0203 | **** | 4.2383 | **** | -1.7704 | **** | 2.1725 | **** | -1.8586 | **** | -4.8444 |
| 167 | $C_2HO$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 168 | $CH_2$ (cyclic) | 0.0478 | 0.0383 | 0.1037 | -0.2351 | 0.0218 | -0.0591 | -0.1159 | -0.1360 | 0.2430 | 0.2778 | 0.2042 | 0.2686 |
| 169 | CH (cyclic) | -0.5705 | -1.3636 | -0.9411 | -0.1917 | -0.3534 | -1.7395 | 0.2311 | -3.3705 | -0.6251 | -0.5402 | -0.1148 | -2.6460 |
| 170 | C (cyclic) | 3.9613 | 1.0459 | 1.1634 | 4.6781 | 0.4554 | 0.8751 | 1.6771 | 2.9940 | -0.9880 | 0.1288 | 0.4143 | -2.4011 |
| 171 | CH=CH (cyclic) | -0.2006 | -0.3717 | -0.1551 | -0.8137 | -0.3694 | -0.4172 | -1.1639 | -0.7469 | 0.0672 | 0.7392 | 0.0478 | 0.6207 |
| 172 | CH=C (cyclic) | 0.0562 | 0.0417 | 0.2281 | -0.4702 | -0.4847 | -0.2261 | -0.1890 | -0.5499 | 0.4875 | -0.5866 | 0.2344 | -0.9710 |
| 173 | C=C (cyclic) | -0.8232 | -3.6378 | -1.2786 | -0.8741 | -2.1582 | -3.0260 | -0.9217 | -3.1614 | -2.1348 | -4.2929 | -1.7673 | -2.0572 |
| 174 | $CH_2$=C (cyclic) | -0.3177 | **** | 0.8420 | **** | -0.3089 | **** | -1.0253 | **** | 1.8053 | **** | 1.4118 | **** |
| 175 | NH (cyclic) | 0.2068 | -0.6063 | -0.2440 | -1.2240 | 0.3465 | -0.7304 | -0.0197 | -0.3171 | 0.2312 | -0.6572 | 0.2839 | -0.4248 |
| 176 | N (cyclic) | 0.7465 | -0.9135 | -0.1568 | -0.0631 | 0.5222 | -1.0567 | 0.5583 | -0.0837 | -0.0920 | -2.0578 | 0.0099 | -1.9278 |
| 177 | CH=N (cyclic) | 0.5309 | 0.0584 | 1.4538 | -0.5310 | 0.6452 | 0.4018 | -0.7708 | -0.1928 | -0.1078 | 1.0813 | 0.1624 | 1.5244 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 178 | C=N (cyclic) | -0.1023 | -0.5296 | -0.0508 | -0.9244 | -0.2505 | -0.4985 | -0.3228 | -0.4970 | -0.0218 | 0.2954 | -0.0959 | 0.3652 |
| 179 | O (cyclic) | -0.6160 | -1.3405 | -0.4177 | -1.5397 | 0.4395 | -1.2790 | -0.1866 | -1.6388 | -0.5459 | -1.2752 | -0.2245 | -0.8744 |
| 180 | CO (cyclic) | -0.1467 | 0.7802 | 0.0273 | 0.3561 | 0.0079 | 0.8369 | 0.6613 | 1.1518 | 0.0450 | 0.9198 | 0.0179 | 0.8191 |
| 181 | S (cyclic) | -0.0113 | -0.4065 | -0.3833 | -0.2744 | 0.3451 | -0.2271 | 0.4112 | -0.2562 | -0.1345 | 0.1483 | 0.0751 | 0.3262 |
| 182 | $SO_2$ (cyclic) | 0.7597 | -0.1628 | -0.2927 | -1.0910 | 1.4921 | -0.9201 | 1.4024 | -0.9510 | 4.2417 | -1.0633 | 4.3097 | -0.8285 |
| 183 | >NH | -0.3552 | -0.1971 | -0.3453 | 0.2549 | -0.1154 | -0.8038 | -0.2881 | -2.1316 | -0.1419 | -1.3337 | -0.1565 | -0.9680 |
| 184 | -O- | 2.1320 | -1.0893 | -0.7573 | -1.8584 | -0.3473 | -1.2017 | -3.3651 | 1.1068 | -1.5205 | -0.8072 | -1.3733 | 0.1945 |
| 185 | -S- | 0.3962 | -0.3454 | 0.8282 | 0.0048 | -1.0937 | -0.6376 | 0.3392 | -0.6979 | 0.5789 | -0.2663 | 0.2048 | -0.1624 |
| 186 | >CO | **** | -0.7055 | **** | -0.5992 | -2.3290 | -0.2531 | -0.3051 | -0.5422 | **** | -0.2774 | -2.3418 | -0.4289 |
| 187 | $PO_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 188 | CH-N | -0.7970 | 0.4190 | -0.7895 | -0.0521 | **** | -1.0914 | **** | 0.6603 | -1.6551 | -1.2596 | **** | -1.9068 |
| 189 | SiHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 190 | SiO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 191 | $SiH_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 192 | SiH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 193 | Si | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 194 | $(CH_3)_3N$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 195 | N=N | **** | -0.6798 | **** | -1.3422 | **** | -0.7632 | **** | -2.2405 | **** | -0.4732 | **** | 0.7231 |
| 196 | $C_{cyclic}$=N- | 0.3309 | -0.3663 | -0.5439 | -1.4243 | 0.1428 | 0.5389 | -0.0601 | 0.0130 | 0.6748 | 0.7104 | 0.3442 | 0.9730 |
| 197 | $C_{cyclic}$=CH- | -1.0346 | **** | -0.2734 | **** | 0.3924 | **** | 0.7763 | **** | -0.4543 | **** | -0.9067 | **** |
| 198 | $C_{cyclic}$=NH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 199 | N=O | -1.8430 | **** | -1.2750 | **** | -1.2403 | **** | -1.3171 | **** | -1.4893 | **** | -1.1820 | **** |
| 200 | $C_{cyclic}$=C | -1.4671 | **** | -0.8366 | **** | -0.0991 | **** | -0.2696 | **** | 0.2699 | **** | -0.5165 | **** |
| 201 | P=O | -5.6926 | -1.8136 | -0.5494 | -0.9626 | -1.0689 | -1.7597 | -0.2499 | -0.7070 | -0.9191 | -1.2194 | -0.9084 | -1.9008 |
| 202 | N=N | 0.5131 | **** | -0.9381 | **** | -0.3220 | **** | -0.1820 | **** | -0.8614 | **** | -0.9633 | **** |
| 203 | C=NH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 204 | >C=S | 0.4140 | -0.4898 | 0.1229 | -0.0780 | -0.0878 | -0.2683 | 0.2252 | -0.1931 | -0.2965 | 0.7406 | -0.1104 | 0.2244 |
| 205 | aC-CON | 0.7937 | -3.7639 | -0.8587 | 0.1618 | -1.0781 | -4.1159 | 1.3443 | **** | -0.6874 | -5.3844 | -0.8192 | -5.1798 |
| 206 | aC=O | 0.6202 | **** | 1.2167 | **** | 1.3160 | **** | 2.2897 | **** | 1.3151 | **** | 1.6017 | **** |
| 207 | aN- | -0.9102 | -1.3674 | -0.4206 | -1.1407 | -0.7226 | -1.4421 | 0.1246 | -0.6597 | -0.2579 | -0.9406 | 0.0202 | -2.6866 |
| 208 | -Na | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 209 | -K | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 210 | HCONH | 0.7064 | **** | 0.1018 | **** | 0.8454 | **** | 0.5271 | **** | 0.4107 | **** | 0.5743 | **** |
| 211 | CHOCH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 212 | $C_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 213 | $SiH_3$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 214 | $SiH_2O$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 215 | CH=C=CH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 216 | CH=C=C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 217 | OP(=S)O | 0.8795 | -0.3584 | 1.6200 | 0.0413 | 0.1379 | -1.0782 | 4.2446 | -0.3807 | -0.0978 | -0.9345 | -0.2608 | -1.2441 |
| 218 | R | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 219 | $CF_{2cyclic}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 220 | $CF_{cyclic}$ | -1.1229 | **** | -1.9644 | **** | -3.4286 | **** | -3.5188 | **** | -3.8755 | **** | -3.4357 | **** |

[a] The symbols $EUA_{C\ 1i}$, $EUA_{NC\ 1i}$, $ERA_{C\ 1i}$, $ERA_{NC\ 1i}$, $EFW_{C\ 1i}$, $EFW_{NC\ 1i}$, $ESW_{C\ 1i}$, $ESW_{NC\ 1i}$, $ENS_{C\ 1i}$, $ENS_{NC\ 1i}$, $EAS_{C\ 1i}$, and $EAS_{NC\ 1i}$ represent the contributions ($C_i$) of the first-order groups for the corresponding properties.

**Table S6. MG Method Based Property Models Analysed Using Simultaneous Regression Method: Second-Order Groups and their Contributions [a] for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), and PCO**

| | Group | $LC50(FM)_{2j}$ | $LC50(DM)_{2j}$ | $LD50_{2j}$ | $LogWs_{2j}$ | $BCF_{2j}$ | $PEL_{2j}$ | $PCO1_{2j}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | $(CH_3)_2CH$ | -0.0418 | 0.2712 | -0.0144 | -0.2681 | 0.3714 | -0.0027 | 0.0313 |
| 2 | $(CH_3)_3C$ | 0.3137 | 4.1440 | 0.1404 | -0.1698 | 0.1995 | 0.0414 | 0.1580 |
| 3 | $CH(CH_3)CH(CH_3)$ | -0.0995 | **** | -0.0161 | -0.3705 | 1.2525 | -0.0350 | 0.0765 |
| 4 | $CH(CH_3)C(CH_3)_2$ | **** | **** | -0.0632 | 0.3614 | 1.1101 | -0.0102 | 0.0354 |
| 5 | $C(CH_3)_2C(CH_3)_2$ | **** | **** | -0.0830 | -0.3041 | **** | -0.0684 | 0.1280 |
| 6 | $CH_n$=$CH_m$-$CH_p$=$CH_k$ (*k,m,n,p* in 0..2) | -0.1772 | 0.8857 | 0.0942 | -0.5589 | -1.2538 | -0.5349 | 0.5521 |
| 7 | $CH_3$-$CH_m$=$CH_n$ (*m,n* in 0..2) | 0.0749 | -3.5587 | -0.1077 | -0.0577 | 0.5100 | -0.1718 | -0.0364 |
| 8 | $CH_2$-$CH_m$=$CH_n$ (*m,n* in 0..2) | -0.2902 | 2.4471 | -0.0401 | 0.0703 | 0.0765 | 0.2436 | 0.0642 |
| 9 | $CH_p$-$CH_m$=$CH_n$ (*m,n* in 0..2; *p* in 0..1) | 0.0477 | -0.5971 | 0.1941 | 0.0841 | -0.0935 | **** | 0.0636 |
| 10 | CHCHO or CCHO | 0.3970 | **** | 0.0140 | -0.5415 | **** | **** | -0.0825 |
| 11 | $CH_3COCH_2$ | 0.0279 | 1.0878 | 0.0083 | 0.4242 | **** | 0.4724 | -0.4764 |
| 12 | $CH_3COCH$ or $CH_3COC$ | 0.5626 | **** | 0.3218 | 0.6789 | -0.3859 | 0.3350 | -0.2513 |
| 13 | CHCOOH or CCOOH | -0.2421 | **** | 0.1954 | 0.2890 | 0.1200 | 0.4992 | -0.1341 |
| 14 | $CH_3COOCH$ or $CH_3COOC$ | 0.0097 | **** | 0.1502 | 0.0716 | **** | -0.0526 | -0.0621 |
| 15 | CO-O-CO | **** | **** | 0.0176 | -0.6806 | 0.0000 | 1.6800 | -10000.0000 |
| 16 | CHOH | -0.0975 | 3.9109 | 0.1155 | 0.2588 | -0.2924 | 0.1197 | 0.0487 |
| 17 | COH | 0.0935 | **** | 0.2278 | 0.8332 | -0.4952 | -0.2124 | 0.4127 |
| 18 | $CH_3COCH_nOH$ (*n* in 0..2) | **** | **** | 0.0054 | 0.9579 | **** | **** | -0.5912 |
| 19 | NCCHOH or NCCOH | **** | **** | 0.2700 | 0.5755 | **** | 0.0000 | -10000.0000 |
| 20 | OH-$CH_n$-COO (*n* in 0..2) | 1.3209 | **** | 0.0627 | 0.1891 | 1.1097 | **** | -0.4055 |
| 21 | $CH_m(OH)CH_n(OH)$ (*m,n* in 0..2) | -1.3676 | 0.0361 | 0.1196 | -0.4326 | -0.7323 | 0.2582 | -0.0929 |
| 22 | $CH_m(OH)CH_n(NH_p)$ (*m,n,p* in 0..2) | -0.4231 | 1.2993 | -0.0387 | 0.1678 | 0.0000 | 0.0860 | -0.0417 |
| 23 | $CH_m(NH_2)CH_n(NH_2)$ (*m,n* in 0..2) | -0.1456 | -3.8727 | -0.5152 | -0.8213 | **** | -1.9649 | **** |
| 24 | $CH_m(NH)CH_n(NH_2)$ (*m,n* in 1..2) | **** | **** | -0.3115 | 0.3349 | 0.0000 | -1.3100 | -10000.0000 |
| 25 | $H_2NCOCH_nCH_mCONH_2$ (*m,n* in 1..2) | **** | **** | **** | -0.7592 | **** | **** | **** |
| 26 | $CH_m(NH_n)$-COOH (*m,n* in 0..2) | 1.2919 | 5.1623 | -0.1667 | -0.4295 | **** | **** | **** |
| 27 | HOOC-$CH_n$-COOH (*n* in 1..2) | **** | **** | -0.0211 | 0.4420 | **** | **** | **** |
| 28 | HOOC-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | 0.1245 | 0.1019 | 3.0655 | **** | -0.4783 |
| 29 | HO-$CH_n$-COOH (*n* in 1..2) | **** | **** | -0.0243 | -0.4726 | 0.8399 | **** | -0.0784 |
| 30 | $NH_2$-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | -0.7592 | **** | **** | **** |
| 31 | $CH_3$-O-$CH_n$-COOH (n in 1..2) | **** | **** | **** | 0.8111 | **** | **** | **** |
| 32 | HS-CH-COOH | 0.0000 | **** | 0.7610 | 0.8831 | -10000.0000 | -1.6100 | -10000.0000 |
| 33 | HS-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | 0.0428 | 1.5568 | **** | **** | **** |
| 34 | NC-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | 0.7607 | 0.5459 | **** | -1.5017 | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 35 | OH-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | 0.3340 | 1.0784 | **** | **** | **** |
| 36 | HS-$CH_n$-$CH_m$-SH (*n, m* in 1..2) | **** | **** | -2.7796 | **** | **** | **** | **** |
| 37 | COO-$CH_n$-$CH_m$-OOC (*n, m* in 1..2) | **** | **** | 0.1410 | -0.4872 | **** | **** | -0.2074 |
| 38 | OOC-$CH_m$-$CH_m$-COO (*n, m* in 1..2) | -0.2344 | -16.0297 | -0.0563 | 0.3977 | 0.0000 | 0.0000 | 0.5500 |
| 39 | NC-$CH_n$-COO (*n* in 1..2) | **** | **** | -0.3421 | -0.2022 | **** | **** | **** |
| 40 | $COCH_n$COO (*n* in 1..2) | **** | **** | -0.0528 | 0.2418 | **** | **** | **** |
| 41 | $CH_m$-O-$CH_n$=$CH_p$ (*m,n,p* in 0..3) | **** | **** | -0.4559 | 0.2041 | 1.6032 | **** | **** |
| 42 | $CH_m$=$CH_n$-F (*m,n* in 0..2) | **** | **** | -0.1574 | 0.0268 | **** | -1.2982 | **** |
| 43 | $CH_m$=$CH_n$-Br (*m,n* in 0..2) | **** | 0.0974 | -0.3767 | 0.4918 | **** | 0.7909 | **** |
| 44 | $CH_m$=$CH_n$-I (*m,n* in 0..2) | **** | **** | **** | 0.0987 | **** | **** | **** |
| 45 | $CH_m$=$CH_n$-Cl (*m,n* in 0..2) | -0.2955 | 0.2971 | 0.1633 | 0.2219 | -0.0569 | -1.7406 | -0.9410 |
| 46 | $CH_m$=$CH_n$-CN (*m,n* in 0..2) | 0.2974 | 4.3101 | 0.2436 | 0.9999 | **** | 0.1129 | 0.0000 |
| 47 | $CH_n$=$CH_m$-COO-$CH_p$ (*m,n,p* in 0..3) | -0.1220 | -0.0227 | -0.1420 | 0.0383 | **** | 0.0206 | 0.0788 |
| 48 | $CH_m$=$CH_n$-CHO (*m,n* in 0..2) | 1.8419 | 9.3811 | 0.1660 | 0.2322 | **** | 0.0772 | 0.3566 |
| 49 | $CH_m$=$CH_n$-COOH (*m,n* in 0..2) | 1.5593 | **** | -0.0256 | -0.0141 | **** | -1.0172 | -0.3647 |
| 50 | aC-$CH_n$-X (*n* in 1..2) X: Halogen | -0.1264 | **** | -0.0660 | 0.1818 | -0.5132 | 0.6493 | **** |
| 51 | aC-$CH_n$-$NH_m$ (*n* in 1..2; *m* in 0..2)) | -0.5979 | **** | -0.2187 | 0.7638 | -0.1295 | **** | **** |
| 52 | aC-$CH_n$-O- (*n* in 1..2) | 0.2287 | -2.8691 | 0.0546 | 0.4210 | 0.1076 | **** | **** |
| 53 | aC-$CH_n$-OH (*n* in 1..2) | 0.0090 | **** | -0.0002 | 0.2106 | -0.3066 | **** | -0.1689 |
| 54 | aC-$CH_n$-CN (*n* in 1..2) | 2.2929 | -2.2058 | 0.0990 | 0.5638 | 0.1066 | **** | **** |
| 55 | aC-$CH_n$-CHO (*n* in 1..2) | **** | **** | -0.3084 | 0.4866 | **** | **** | **** |
| 56 | aC-$CH_n$-SH (*n* in 1..2) | **** | **** | 0.6393 | **** | **** | **** | **** |
| 57 | aC-$CH_n$-COOH (*n* in 1..2) | **** | **** | 0.3223 | 0.2576 | -0.6363 | **** | **** |
| 58 | aC-$CH_n$-CO- (*n* in 1..2) | 0.0000 | **** | 0.1190 | 0.3883 | -0.3370 | -10000.0000 | -10000.0000 |
| 59 | aC-$CH_n$-S- (*n* in 1..2) | **** | **** | -0.5222 | 0.1038 | 1.1115 | **** | **** |
| 60 | aC-$CH_n$-OOC-H (*n* in 1..2) | **** | **** | 0.2836 | **** | **** | **** | **** |
| 61 | aC-$CH_m$-$NO_2$ (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** |
| 62 | aC-$CH_n$-$CONH_2$ (*n* in 1..2) | **** | **** | 0.1814 | 0.5403 | **** | **** | **** |
| 63 | aC-$CH_n$-OOC (*n* in 1..2) | -0.4720 | -0.4810 | -0.1659 | -0.7164 | -0.3023 | 0.7772 | **** |
| 64 | aC-$CH_n$-COO (*n* in 1..2) | -0.6542 | -6.1400 | 0.0539 | -0.1301 | 0.0458 | **** | **** |
| 65 | aC-$SO_2$-OH | -0.5382 | **** | -0.1870 | 1.0300 | 1.5586 | **** | **** |
| 66 | aC-CH($CH_3$)$_2$ | 0.0851 | -1.2681 | -0.3378 | -0.2936 | 0.1272 | -1.1271 | 0.0187 |
| 67 | aC-C($CH_3$)$_3$ | 0.9465 | **** | -0.1434 | 0.7323 | -0.0775 | 5.6444 | 0.0907 |
| 68 | aC-$CF_3$ | 0.1166 | 0.3533 | -0.7788 | 1.9593 | 1.6118 | **** | -0.2101 |
| 69 | ($CH_n$=C)$_{cyc}$-CHO (*n* in 0..2) | **** | **** | -0.1031 | -0.5395 | **** | 0.2158 | **** |
| 70 | ($CH_n$=C)$_{cyc}$-COO-$CH_m$ (*n,m* in 0..3) | **** | **** | -0.2405 | -0.0249 | 0.7906 | **** | **** |
| 71 | ($CH_n$=C)$_{cyc}$-CO- (*n* in 0..2) | -0.6127 | **** | 0.1205 | -0.2214 | **** | **** | **** |
| 72 | ($CH_n$=C)$_{cyc}$-$CH_3$ (*n* in 0..2) | -0.3382 | 5.1943 | -0.1284 | 0.1409 | -0.2985 | -2.5355 | -0.2129 |
| 73 | ($CH_n$=C)$_{cyc}$-$CH_2$ (*n* in 0..2) | -0.8382 | -2.2216 | -0.0219 | 0.4412 | **** | 0.1846 | **** |
| 74 | ($CH_n$=C)$_{cyc}$-CN (*n* in 0..2) | -0.0901 | **** | -0.6381 | -0.3809 | **** | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 75 | $(CH_n{=}C)_{cyc}$-Cl (*n* in 0..2) | -0.2396 | 0.4435 | 0.0291 | 0.1416 | -0.3723 | -0.7029 | **** |
| 76 | $CH_{cyc}$-$CH_3$ | 0.0321 | 0.7237 | -0.2120 | 0.1410 | 0.1321 | -0.3707 | -0.0854 |
| 77 | $CH_{cyc}$-$CH_2$ | -0.5631 | 1.4904 | -0.2032 | 0.3006 | 0.7223 | -1.1768 | -0.0314 |
| 78 | $CH_{cyc}$-CH | 0.6735 | -0.5530 | -0.2439 | 0.3390 | -0.0178 | **** | 0.0269 |
| 79 | $CH_{cyc}$-C | -0.6977 | **** | -0.4669 | 0.0614 | 1.3638 | **** | -0.0579 |
| 80 | $CH_{cyc}$-CH=$CH_n$ (*n* in 1..2) | 0.0547 | -3.3258 | -0.0845 | -0.0140 | 0.4312 | 1.4915 | **** |
| 81 | $CH_{cyc}$-C=$CH_n$ (*n* in 1..2) | -0.6174 | 2.9140 | -0.4441 | 0.5030 | -2.8314 | **** | 0.8562 |
| 82 | $CH_{cyc}$-Cl | -0.3183 | -0.9800 | -0.0649 | 0.2801 | 0.2869 | -1.2358 | **** |
| 83 | $CH_{cyc}$-F | **** | **** | **** | 0.5092 | **** | **** | **** |
| 84 | $CH_{cyc}$-OH | -0.2028 | 1.0683 | -0.1594 | -0.0298 | -0.2648 | 0.0492 | -0.1266 |
| 85 | $CH_{cyc}$-$NH_2$ | **** | **** | -0.4730 | 0.1626 | **** | 0.2963 | **** |
| 86 | $CH_{cyc}$-NH-$CH_n$ (*n* in 0..3) | 0.1644 | -0.7336 | -0.3636 | 0.6737 | 0.0436 | **** | **** |
| 87 | $CH_{cyc}$-N-$CH_n$ (*n* in 0..3) | 1.2144 | **** | -0.4944 | 0.6127 | 0.8210 | **** | **** |
| 88 | $CH_{cyc}$-SH | **** | **** | 1.6279 | **** | **** | 0.9254 | **** |
| 89 | $CH_{cyc}$-CN | **** | **** | -0.5241 | 0.9679 | **** | **** | **** |
| 90 | $CH_{cyc}$-COOH | 0.3786 | **** | -0.3532 | 0.0041 | 0.9628 | **** | **** |
| 91 | $CH_{cyc}$-CO | **** | **** | -0.0387 | 0.4932 | **** | **** | **** |
| 92 | $CH_{cyc}$-$NO_2$ | **** | **** | 0.4482 | 0.9901 | **** | **** | **** |
| 93 | $CH_{cyc}$-S- | -1.6152 | **** | -0.2856 | 0.8968 | **** | **** | **** |
| 94 | $CH_{cyc}$-CHO | 0.5061 | **** | -0.3566 | -0.0628 | **** | **** | -0.0370 |
| 95 | $CH_{cyc}$-O- | **** | **** | -0.4260 | 0.2118 | **** | 2.1204 | **** |
| 96 | $CH_{cyc}$-OOCH | **** | **** | 0.0509 | **** | **** | **** | **** |
| 97 | $CH_{cyc}$-COO | -0.7603 | -1.1127 | -0.4650 | -0.0666 | 0.2877 | **** | **** |
| 98 | $CH_{cyc}$-OOC | -0.4578 | -2.0921 | -0.3600 | 0.2191 | **** | **** | **** |
| 99 | $C_{cyc}$-$CH_3$ | 0.5272 | -0.6499 | -0.1205 | -0.0642 | -0.0631 | 2.4390 | -0.0002 |
| 100 | $C_{cyc}$-$CH_2$ | 1.2910 | -0.7346 | -0.0003 | 0.4062 | -0.5229 | 1.0713 | -0.0319 |
| 101 | $C_{cyc}$-OH | -0.2373 | -1.9432 | 0.2718 | -0.3541 | -1.6360 | -1.1715 | **** |
| 102 | >$N_{cyc}$-$CH_3$ | 0.5632 | 6.9066 | 0.0506 | 0.4460 | -1.2764 | **** | -0.2283 |
| 103 | >$N_{cyc}$-$CH_2$ | -0.3114 | 1.9548 | 0.0864 | 0.5720 | -0.5023 | **** | **** |
| 104 | AROMRINGs$^1$s$^2$ | 0.2621 | 0.2164 | 0.0716 | -0.1196 | 0.1105 | -0.4639 | -0.2439 |
| 105 | AROMRINGs$^1$s$^3$ | -0.1180 | 0.1443 | 0.0941 | -0.2204 | 0.0452 | -0.5725 | -0.2979 |
| 106 | AROMRINGs$^1$s$^4$ | 0.2146 | 0.3769 | 0.0225 | -0.1809 | -0.0630 | 0.0847 | -0.1949 |
| 107 | AROMRINGs$^1$s$^2$s$^3$ | -0.0720 | 0.3149 | 0.1488 | 0.0218 | 0.3649 | -0.6955 | -0.2645 |
| 108 | AROMRINGs$^1$s$^2$s$^4$ | 0.2048 | 0.0338 | 0.0290 | -0.0516 | 0.1049 | -0.9649 | -0.1603 |
| 109 | AROMRINGs$^1$s$^3$s$^5$ | -0.0500 | -0.1186 | 0.0747 | -0.3272 | 0.4939 | -0.7364 | -0.2885 |
| 110 | AROMRINGs$^1$s$^2$s$^3$s$^4$ | 0.4147 | -0.4579 | -0.1170 | 0.1183 | 0.0951 | **** | -0.1422 |
| 111 | AROMRINGs$^1$s$^2$s$^3$s$^5$ | -0.0011 | -0.0506 | -0.0488 | 0.0218 | 0.2253 | -3.3441 | -0.0812 |
| 112 | AROMRINGs$^1$s$^2$s$^4$s$^5$ | 0.3494 | -0.5777 | 0.0164 | -0.2460 | 0.5129 | **** | -0.1437 |
| 113 | PYRIDINEs$^2$ | -0.9133 | **** | 0.4753 | 0.9191 | -0.1885 | **** | **** |
| 114 | PYRIDINEs$^3$ | -0.4916 | **** | -0.0360 | 0.9089 | -1.0214 | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 115 | PYRIDINEs$^{4}$ | -0.7575 | **** | 0.2708 | 0.4137 | -0.6056 | **** | **** |
| 116 | PYRIDINEs$^{2}$s$^{3}$ | -1.1257 | **** | 0.5297 | 0.4975 | **** | **** | **** |
| 117 | PYRIDINEs$^{2}$s$^{4}$ | **** | **** | 0.2268 | 0.2720 | **** | **** | **** |
| 118 | PYRIDINEs$^{2}$s$^{5}$ | -0.6158 | 0.2969 | 0.0882 | 0.2547 | 1.5410 | **** | **** |
| 119 | PYRIDINEs$^{2}$s$^{6}$ | -0.2151 | **** | 0.3627 | -0.0288 | **** | **** | **** |
| 120 | PYRIDINEs$^{3}$s$^{4}$ | **** | **** | 0.2749 | 0.1079 | **** | **** | **** |
| 121 | PYRIDINEs$^{3}$s$^{5}$ | **** | **** | -0.2183 | -0.1755 | **** | **** | **** |
| 122 | PYRIDINEs$^{2}$s$^{3}$s$^{6}$ | **** | **** | -0.5987 | -0.2032 | **** | **** | **** |
| 123 | (CH$^{n}$=CH$^{m}$)$^{cyc}$-COOH | **** | **** | -0.9943 | -0.5168 | **** | **** | **** |
| 124 | AROMRINGs$^{1}$s$^{2}$s$^{3}$s$^{4}$s$^{5}$ | -0.0481 | -0.6732 | 0.1845 | 0.3517 | 0.1277 | **** | 0.0474 |
| 125 | aC-NHCOCH$_2$N | **** | **** | 0.6326 | 1.1063 | **** | **** | **** |
| 126 | (N=C)$_{cyc}$-CH$_3$ | -0.5150 | **** | 0.0283 | 0.6955 | -2.0489 | **** | **** |
| 127 | aC-CONH(CH$_2$)$_2$N | **** | **** | 0.2557 | 0.7687 | **** | **** | **** |
| 128 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | **** | **** | -0.1618 | 0.2409 | 0.7943 | **** | **** |
| 129 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | **** | **** | 0.2385 | 0.8555 | **** | **** | **** |
| 130 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | **** | **** | -0.3188 | 0.2319 | 1.6555 | **** | **** |

[a] The symbols LC50(FM)$_{2j}$, LC50(DM)$_{2j}$LD50$_{2j}$, LogWs$_{2j}$, BCF$_{2j,}$ PEL$_{2j,}$ PCO$_{2j}$ represent the contributions ($D_j$) of the second-order groups for the corresponding properties. Note that there are no second-order contributions for GWP, ODP and AP since these properties belong to chemicals whose structure is defined by first-order groups alone.

**Table S6 (continued). MG Method Based Property Models Analysed Using <u>Simultaneous Regression Method</u>: Second-Order Groups and their Contributions [a] for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| | Group | $EUA_{C\,2j}$ | $EUA_{NC\,2j}$ | $ERA_{C\,2j}$ | $ERA_{NC2j}$ | $EFW_{C\,2j}$ | $EFW_{NC\,2j}$ | $ESW_{C\,2j}$ | $ESW_{NC\,2j}$ | $ENS_{C\,2j}$ | $ENS_{NC\,2j}$ | $EAS_{C\,2j}$ | $EAS_{NC\,2j}$ | $EUA_{C\,2j}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $(CH_3)_2CH$ | 0.4454 | 0.7777 | 0.0850 | 0.9533 | -0.0271 | 0.6037 | 0.3571 | 1.0021 | -0.1402 | 0.9090 | 0.2321 | 1.0531 | 0.4454 |
| 2 | $(CH_3)_3C$ | -1.0261 | 0.7628 | -0.3870 | 0.2392 | -0.2564 | 0.6369 | -1.2769 | 1.2133 | -0.8796 | 1.1569 | 0.1557 | 0.4723 | -1.0261 |
| 3 | $CH(CH_3)CH(CH_3)$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 4 | $CH(CH_3)C(CH_3)_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 5 | $C(CH_3)_2C(CH_3)_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 6 | $CH_n$=$CH_m$-$CH_p$=$CH_k$ (*k,m,n,p* in 0..2) | 1.7943 | 2.7047 | -0.9739 | 0.2460 | -0.9041 | 1.5568 | -0.7469 | 1.3600 | -0.2300 | 2.2724 | -0.4340 | 0.4036 | 1.7943 |
| 7 | $CH_3$-$CH_m$=$CH_n$ (*m,n* in 0..2) | 0.8992 | 1.1266 | -0.0583 | 0.2893 | -0.1388 | 0.5668 | 0.0073 | 0.1956 | -0.1366 | 1.3767 | -0.0679 | 0.5363 | 0.8992 |
| 8 | $CH_2$-$CH_m$=$CH_n$ (*m,n* in 0..2) | 0.6538 | 0.2386 | -0.3203 | 1.2598 | -0.4937 | 0.5536 | -0.9016 | -0.0152 | -0.3931 | 0.7228 | -0.6658 | 0.1489 | 0.6538 |
| 9 | $CH_p$-$CH_m$=$CH_n$ (*m,n* in 0..2; *p* in 0..1) | -0.1003 | -1.0487 | -1.2161 | - | -0.9001 | -1.3775 | -1.0900 | -1.1375 | -1.0261 | -1.2026 | -1.3403 | -1.3735 | -0.1003 |
| 10 | CHCHO or CCHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | $CH_3COCH_2$ | **** | -0.7929 | **** | 0.0988 | 0.2147 | -0.4836 | **** | -0.2261 | **** | -0.4645 | 0.0150 | -0.6226 | **** |
| 12 | $CH_3COCH$ or $CH_3COC$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | CHCOOH or CCOOH | -0.3208 | 0.0291 | -1.4573 | - | -1.4573 | -0.3280 | -1.5177 | -0.5655 | -1.2028 | -0.7276 | -0.8287 | -0.2615 | -0.3208 |
| 14 | $CH_3COOCH$ or $CH_3COOC$ | -0.9094 | **** | -0.1570 | **** | -0.3162 | **** | -1.2374 | **** | -1.1378 | **** | -0.5791 | **** | -0.9094 |
| 15 | CO-O-CO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 16 | CHOH | -0.5119 | -0.8836 | -0.6160 | 0.1893 | -1.0520 | -0.8619 | -0.6700 | -0.9111 | -0.4799 | -0.2670 | -0.5091 | -0.3797 | -0.5119 |
| 17 | COH | -0.3245 | -2.3976 | -1.5910 | - | -1.7123 | -1.5682 | -1.1338 | -1.8517 | -1.0148 | -1.1666 | -1.2208 | -1.0103 | -0.3245 |
| 18 | $CH_3COCH_nOH$ (*n* in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 19 | NCCHOH or NCCOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 20 | OH-$CH_n$-COO (*n* in 0..2) | -0.2640 | -2.5789 | -1.3986 | - | -4.3790 | -4.1296 | -1.7941 | -4.8709 | -3.8361 | 1.2855 | -2.3116 | 1.1969 | -0.2640 |
| 21 | $CH_m(OH)CH_n(OH)$ (*m,n* in 0..2) | -2.6993 | **** | **** | - | **** | **** | **** | **** | **** | **** | **** | **** | -2.6993 |
| 22 | $CH_m(OH)CH_n(NH_p)$ (*m,n,p* in 0..2) | -1.6155 | **** | -1.8221 | **** | -2.0086 | **** | -2.5453 | **** | -2.2315 | **** | -2.2681 | **** | -1.6155 |
| 23 | $CH_m(NH_2)CH_n(NH_2)$ (*m,n* in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 24 | $CH_m(NH)CH_n(NH_2)$ (*m,n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 25 | $H_2NCOCH_nCH_mCONH_2$ (*m,n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 26 | $CH_m(NH_n)$-COOH (*m,n* in 0..2) | -0.1495 | **** | -0.3633 | **** | -0.4518 | 0.7218 | -0.6857 | **** | -0.1520 | 0.3571 | 0.1525 | 1.3200 | -0.1495 |
| 27 | HOOC-$CH_n$-COOH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 28 | HOOC-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 29 | HO-$CH_n$-COOH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 30 | $NH_2$-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 31 | $CH_3$-O-$CH_n$-COOH (n in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 32 | HS-CH-COOH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | HS-$CH_n$-$CH_m$-COOH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 34 | NC-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 35 | OH-$CH_n$-$CH_m$-CN (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 36 | HS-$CH_n$-$CH_m$-SH (*n, m* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 37 | COO-$CH_n$-$CH_m$-OOC (*n, m* in 1..2) | 0.4211 | **** | 1.1314 | **** | 1.2644 | **** | -0.3272 | **** | 1.5168 | **** | 0.9057 | **** | 0.4211 |
| 38 | OOC-$CH_m$-$CH_m$-COO (*n, m* in 1..2) | -1.5241 | 3.6672 | -0.9983 | 3.6117 | -0.6690 | 4.6673 | -1.0778 | 5.4329 | 1.4117 | 4.1810 | 0.0326 | 3.9405 | -1.5241 |
| 39 | NC-$CH_n$-COO (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 40 | $COCH_nCOO$ (*n* in 1..2) | -2.5340 | **** | -2.0342 | **** | -2.6213 | **** | -4.7273 | **** | 0.0760 | **** | -1.0854 | **** | -2.5340 |
| 41 | $CH_m$-O-$CH_n$=$CH_p$ (*m,n,p* in 0..3) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 42 | $CH_m$=$CH_n$-F (*m,n* in 0..2) | 0.3786 | **** | 0.1935 | **** | -0.0139 | **** | 1.4655 | **** | 0.6120 | **** | 0.3624 | **** | 0.3786 |
| 43 | $CH_m$=$CH_n$-Br (*m,n* in 0..2) | **** | 0.6806 | **** | 0.5414 | **** | 0.2753 | **** | 0.0771 | **** | 0.9150 | **** | 0.3664 | **** |
| 44 | $CH_m$=$CH_n$-I (*m,n* in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 45 | $CH_m$=$CH_n$-Cl (*m,n* in 0..2) | 2.8096 | 0.4492 | 0.1697 | 1.2437 | 0.9242 | 0.0103 | 0.5423 | 0.0245 | 0.3108 | 0.6621 | 0.9509 | -0.0168 | 2.8096 |
| 46 | $CH_m$=$CH_n$-CN (*m,n* in 0..2) | -0.2510 | -0.7132 | -0.6328 | - | -1.2833 | -0.4240 | -2.0346 | -1.5374 | -1.6324 | -0.4531 | -1.7620 | -1.4200 | -0.2510 |
| 47 | $CH_n$=$CH_m$-COO-$CH_p$ (*m,n,p* in 0..3) | 0.4041 | 1.7879 | -0.9669 | 0.9356 | -0.5581 | 1.7600 | 0.4096 | 1.5587 | -0.5862 | 1.3524 | -0.4495 | 0.7187 | 0.4041 |
| 48 | $CH_m$=$CH_n$-CHO (*m,n* in 0..2) | -0.2428 | -1.7213 | -1.0203 | - | -0.8167 | -0.6646 | 0.0942 | -3.1005 | -0.9783 | -1.7167 | -1.0652 | -1.9796 | -0.2428 |
| 49 | $CH_m$=$CH_n$-COOH (*m,n* in 0..2) | -1.0794 | **** | -2.7056 | **** | -2.6653 | **** | -2.8203 | **** | -2.2235 | **** | -1.9522 | **** | -1.0794 |
| 50 | aC-$CH_n$-X (*n* in 1..2) X: Halogen | 2.3361 | **** | 0.8709 | **** | 1.3713 | **** | 0.9893 | **** | 0.6633 | **** | 1.2260 | **** | 2.3361 |
| 51 | aC-$CH_n$-$NH_m$ (*n* in 1..2; *m* in 0..2)) | 1.0687 | **** | 0.6693 | **** | 0.6495 | **** | 0.5118 | **** | 0.9552 | **** | 0.9981 | **** | 1.0687 |
| 52 | aC-$CH_n$-O- (*n* in 1..2) | 6.0723 | 0.8650 | 0.7509 | 0.8463 | 2.0927 | 0.8340 | 8.1599 | 0.5995 | 3.2998 | 1.1806 | 2.5796 | 1.3480 | 6.0723 |
| 53 | aC-$CH_n$-OH (*n* in 1..2) | 0.7158 | -1.0641 | 1.2554 | - | 1.2648 | -1.0102 | 1.4622 | -0.7395 | 1.1325 | -1.3372 | 1.0486 | -1.3613 | 0.7158 |
| 54 | aC-$CH_n$-CN (*n* in 1..2) | **** | 0.6234 | **** | 0.7483 | **** | 0.5938 | **** | 0.8119 | **** | 0.1023 | **** | -0.5529 | **** |
| 55 | aC-$CH_n$-CHO (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 56 | aC-$CH_n$-SH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 57 | aC-$CH_n$-COOH (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 58 | aC-$CH_n$-CO- (*n* in 1..2) | **** | -0.4253 | **** | - | **** | 0.0434 | **** | 0.2368 | **** | -0.7374 | **** | -0.4952 | **** |
| 59 | aC-$CH_n$-S- (*n* in 1..2) | **** | -0.4363 | **** | - | **** | 0.1549 | **** | 0.2885 | **** | 1.0150 | **** | 0.9068 | **** |
| 60 | aC-$CH_n$-OOC-H (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 61 | aC-$CH_m$-$NO_2$ (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 62 | aC-$CH_n$-$CONH_2$ (*n* in 1..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 63 | aC-$CH_n$-OOC (*n* in 1..2) | 2.1857 | -0.8637 | 1.5944 | - | 1.7073 | -0.2397 | 2.2699 | -0.3791 | 1.8470 | -0.4877 | 1.5479 | -0.5280 | 2.1857 |
| 64 | aC-$CH_n$-COO (*n* in 1..2) | 0.6946 | 2.4109 | 1.1951 | 1.7731 | 3.0189 | 3.2424 | 1.4124 | 3.0979 | 2.4991 | 0.0813 | 1.5682 | -0.2439 | 0.6946 |
| 65 | aC-$SO_2$-OH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 66 | aC-$CH(CH_3)_2$ | -0.6750 | 0.3543 | -1.5063 | - | -0.9091 | 0.7285 | 0.6240 | 1.5970 | -0.6537 | -0.2616 | -0.4466 | 0.4541 | -0.6750 |
| 67 | aC-$C(CH_3)_3$ | 3.0869 | 1.3842 | 2.6282 | **** | 3.8946 | **** | 3.3901 | **** | 3.7734 | 3.3323 | 3.7719 | **** | 3.0869 |
| 68 | aC-$CF_3$ | 2.7512 | 1.2533 | 5.7320 | 4.7670 | 4.2528 | 1.3657 | 8.4099 | 1.9079 | 5.3555 | 1.0867 | 4.3307 | 1.3526 | 2.7512 |
| 69 | $(CH_n=C)_{cyc}$-CHO (*n* in 0..2) | 2.1717 | 1.2393 | 0.8882 | 1.6531 | 0.8121 | 0.8593 | 1.6104 | 2.0521 | 1.2094 | 0.5196 | 1.0755 | 0.8454 | 2.1717 |
| 70 | $(CH_n=C)_{cyc}$-COO-$CH_m$ (*n,m* in 0..3) | **** | 6.2390 | **** | 2.1891 | **** | 4.7572 | **** | 5.4373 | **** | 2.2894 | **** | -0.4511 | **** |
| 71 | $(CH_n=C)_{cyc}$-CO- (*n* in 0..2) | **** | 3.6588 | **** | 1.5130 | **** | 1.7279 | **** | 2.6167 | **** | 1.7446 | **** | 0.2172 | **** |
| 72 | $(CH_n=C)_{cyc}$-$CH_3$ (*n* in 0..2) | 0.0144 | 1.6249 | 0.2798 | 0.8405 | 0.6716 | 1.7910 | -0.2938 | 2.6849 | -0.1270 | 2.3612 | 0.1740 | 1.0475 | 0.0144 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 73 | $(CH_n{=}C)_{cyc}$-$CH_2$ ($n$ in 0..2) | 0.8733 | -0.4514 | 0.1148 | 0.5525 | 0.0473 | -0.3333 | 0.5312 | -0.0406 | -0.5548 | 1.1360 | -0.4881 | 0.8778 | 0.8733 |
| 74 | $(CH_n{=}C)_{cyc}$-CN ($n$ in 0..2) | **** | 2.8009 | **** | 1.0755 | **** | 1.6933 | **** | 1.5060 | **** | 2.8493 | **** | 0.0293 | **** |
| 75 | $(CH_n{=}C)_{cyc}$-Cl ($n$ in 0..2) | 2.4853 | 0.8650 | 1.3227 | 1.0251 | 1.5691 | 0.5254 | 1.5760 | 1.1605 | 1.3515 | 1.0003 | 1.6352 | 0.3695 | 2.4853 |
| 76 | $CH_{cyc}$-$CH_3$ | 1.3286 | 3.5106 | 0.6868 | 1.6797 | -0.4296 | 3.1837 | -0.2258 | 5.4130 | 0.1520 | 2.6395 | -0.3265 | 3.2632 | 1.3286 |
| 77 | $CH_{cyc}$-$CH_2$ | 0.8847 | 3.8741 | 1.1317 | 2.1603 | -0.3996 | 3.4800 | -0.4124 | 6.1122 | 0.0876 | 3.1659 | -0.5379 | 3.5261 | 0.8847 |
| 78 | $CH_{cyc}$-CH | **** | -0.9238 | **** | 2.3200 | **** | 0.7457 | **** | 6.1796 | **** | 0.7722 | **** | 2.8086 | **** |
| 79 | $CH_{cyc}$-C | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 80 | $CH_{cyc}$-$CH{=}CH_n$ ($n$ in 1..2) | 2.3806 | 1.0425 | 1.8181 | 2.9808 | 0.9041 | 1.1953 | 0.2788 | 3.5141 | 1.3798 | 1.1572 | 0.2457 | 3.8725 | 2.3806 |
| 81 | $CH_{cyc}$-$C{=}CH_n$ ($n$ in 1..2) | 1.1720 | 2.4730 | **** | 2.6495 | -1.7594 | 2.6481 | -0.9170 | 3.5831 | **** | 0.6485 | -2.9697 | 0.2060 | 1.1720 |
| 82 | $CH_{cyc}$-Cl | 2.3759 | 0.9708 | 1.2937 | 1.3585 | 0.5164 | 1.1560 | 0.5041 | 3.6834 | 0.4250 | -0.1607 | 0.4479 | 1.8211 | 2.3759 |
| 83 | $CH_{cyc}$-F | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 84 | $CH_{cyc}$-OH | -1.3384 | **** | 1.3004 | **** | 2.3858 | **** | -0.9943 | **** | 2.7345 | **** | 0.7397 | **** | -1.3384 |
| 85 | $CH_{cyc}$-$NH_2$ | **** | 1.6149 | **** | 1.4558 | **** | 2.6148 | **** | 3.5476 | **** | 0.1803 | **** | 2.2142 | **** |
| 86 | $CH_{cyc}$-NH-$CH_n$ ($n$ in 0..3) | **** | 0.8163 | **** | - | **** | 1.3080 | **** | 3.3778 | **** | 0.6604 | **** | 2.4904 | **** |
| 87 | $CH_{cyc}$-N-$CH_n$ ($n$ in 0..3) | **** | **** | **** | **** | **** | **** | **** | **** | 1.1339 | **** | **** | **** | **** |
| 88 | $CH_{cyc}$-SH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 89 | $CH_{cyc}$-CN | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 90 | $CH_{cyc}$-COOH | -0.1192 | **** | -0.0258 | **** | -0.8363 | **** | -0.2251 | **** | 0.1528 | **** | -0.3452 | **** | -0.1192 |
| 91 | $CH_{cyc}$-CO | -2.0389 | **** | -0.4304 | **** | 0.8105 | **** | 0.8953 | **** | 0.5810 | **** | -0.5264 | **** | -2.0389 |
| 92 | $CH_{cyc}$-$NO_2$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 93 | $CH_{cyc}$-S- | **** | 3.3988 | **** | 2.0753 | **** | 4.0643 | **** | 6.2953 | **** | 1.4916 | **** | 3.3534 | **** |
| 94 | $CH_{cyc}$-CHO | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 95 | $CH_{cyc}$-O- | -2.2400 | 1.6824 | -0.1341 | 1.9541 | -0.6127 | 2.8569 | -1.1417 | 3.6426 | -0.9768 | -0.4946 | -1.8716 | 2.9381 | -2.2400 |
| 96 | $CH_{cyc}$-OOCH | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 97 | $CH_{cyc}$-COO | 1.1500 | 5.8631 | 2.5531 | - | 1.0973 | 7.4198 | 3.3882 | 7.2877 | 3.2162 | 3.5543 | 3.4197 | 2.2421 | 1.1500 |
| 98 | $CH_{cyc}$-OOC | -2.6029 | **** | -1.7657 | **** | 0.5728 | **** | -2.1338 | **** | -1.5802 | **** | -2.7917 | **** | -2.6029 |
| 99 | $C_{cyc}$-$CH_3$ | -1.2509 | -0.7606 | -0.4669 | - | -0.3483 | -0.7903 | -0.0784 | -1.1166 | 0.3823 | -0.8171 | -0.2401 | 1.2917 | -1.2509 |
| 100 | $C_{cyc}$-$CH_2$ | -2.1932 | -1.6180 | -1.4721 | - | -1.1779 | -1.6672 | -1.1572 | -3.6152 | -0.5213 | -0.5739 | -1.0688 | 0.5956 | -2.1932 |
| 101 | $C_{cyc}$-OH | -4.1529 | **** | -2.6517 | **** | -2.0853 | **** | -2.1961 | **** | -1.3210 | **** | -2.4397 | **** | -4.1529 |
| 102 | $>N_{cyc}$-$CH_3$ | -2.6236 | 1.2241 | -0.6600 | 1.0941 | -1.6712 | 1.8657 | -0.3848 | 2.4545 | -0.5769 | 1.1601 | -1.1689 | 3.0465 | -2.6236 |
| 103 | $>N_{cyc}$-$CH_2$ | -2.3077 | 0.4390 | -1.6670 | 0.3304 | -2.0988 | 0.3703 | -5.2922 | 0.7772 | -1.0751 | 0.4878 | -1.1755 | -0.0732 | -2.3077 |
| 104 | AROMRINGs$^1$s$^2$ | -0.2932 | 0.3709 | -0.7671 | 0.2911 | -0.5028 | 0.2541 | 0.0123 | 0.0459 | -0.5034 | 0.0076 | -0.6684 | 0.3018 | -0.2932 |
| 105 | AROMRINGs$^1$s$^3$ | -0.0386 | -0.2628 | 0.2833 | - | 0.1137 | -0.7375 | -0.3048 | -0.8663 | -0.4912 | -0.4973 | -0.2536 | -0.6056 | -0.0386 |
| 106 | AROMRINGs$^1$s$^4$ | 0.2368 | -0.2471 | -0.5227 | - | -0.2714 | 0.0439 | 0.2671 | 0.0037 | -0.5270 | -0.1168 | -0.6095 | 0.1734 | 0.2368 |
| 107 | AROMRINGs$^1$s$^2$s$^3$ | -0.8549 | -0.8951 | -1.4756 | - | -1.5103 | -1.2564 | -1.0988 | -1.9660 | -1.7515 | -1.0944 | -1.9131 | -1.2851 | -0.8549 |
| 108 | AROMRINGs$^1$s$^2$s$^4$ | 0.2638 | 0.1433 | -0.2493 | - | -0.0359 | -0.4147 | 0.5094 | -0.8571 | -0.3209 | -0.6798 | -0.6706 | -0.5102 | 0.2638 |
| 109 | AROMRINGs$^1$s$^3$s$^5$ | 0.9570 | 1.0957 | -0.1781 | - | -0.0445 | 1.1302 | 0.7663 | -1.6879 | 0.2129 | 1.4766 | -0.2338 | 1.4124 | 0.9570 |
| 110 | AROMRINGs$^1$s$^2$s$^3$s$^4$ | -0.4034 | 0.1981 | -0.6652 | 0.4054 | -0.7230 | -0.1871 | -0.0139 | -1.7111 | -0.7600 | 0.2076 | -1.0221 | 0.1758 | -0.4034 |
| 111 | AROMRINGs$^1$s$^2$s$^3$s$^5$ | 0.2665 | -0.0685 | 0.1411 | - | 0.2443 | -0.2365 | 0.7054 | -0.8478 | 0.3094 | -0.1337 | -0.0754 | -0.1363 | 0.2665 |
| 112 | AROMRINGs$^1$s$^2$s$^4$s$^5$ | 1.4506 | 0.1108 | 0.0440 | 0.1174 | -0.3537 | 0.1461 | 0.3393 | 0.0988 | -0.6499 | -0.2285 | -0.8629 | 0.4303 | 1.4506 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 113 | PYRIDINEs$^{2}$ | -3.9107 | **** | -1.3864 | **** | -1.3452 | **** | -5.7261 | **** | -0.5775 | **** | -1.1313 | **** | -3.9107 |
| 114 | PYRIDINEs$^{3}$ | 3.9544 | **** | 2.0083 | **** | 1.0925 | **** | 2.1646 | **** | 1.8698 | **** | 1.7514 | **** | 3.9544 |
| 115 | PYRIDINEs$^{4}$ | 3.7069 | **** | 1.5478 | **** | 0.3720 | **** | 1.3554 | **** | 1.2727 | **** | 0.7867 | **** | 3.7069 |
| 116 | PYRIDINEs$^{2}$s$^{3}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 117 | PYRIDINEs$^{2}$s$^{4}$ | 1.2066 | **** | 0.9353 | **** | 1.1112 | **** | 2.3016 | **** | 1.0532 | **** | 0.9182 | **** | 1.2066 |
| 118 | PYRIDINEs$^{2}$s$^{5}$ | **** | -0.7277 | **** | - | **** | 0.3373 | **** | 1.9774 | **** | 0.4128 | **** | 0.5314 | **** |
| 119 | PYRIDINEs$^{2}$s$^{6}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 120 | PYRIDINEs$^{3}$s$^{4}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 121 | PYRIDINEs$^{3}$s$^{5}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 122 | PYRIDINEs$^{2}$s$^{3}$s$^{6}$ | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 123 | (CH$^{n}$=CH$^{m}$)$^{cyc}$-COOH | 1.6652 | **** | 0.5202 | **** | 0.8885 | **** | 4.8555 | **** | -1.0278 | **** | -0.2002 | **** | 1.6652 |
| 124 | AROMRINGs$^{1}$s$^{2}$s$^{3}$s$^{4}$s$^{5}$ | -1.8147 | -1.2355 | -2.6357 | - | -1.7893 | 0.0961 | -2.1346 | -0.4804 | -5.3012 | -0.1426 | -3.6115 | -0.8818 | -1.8147 |
| 125 | aC-NHCOCH$_2$N | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 126 | (N=C)$_{cyc}$-CH$_3$ | 1.7027 | -2.5588 | 0.5750 | 3.4910 | 1.1188 | -1.0985 | 1.9860 | -1.5018 | 0.4178 | -0.1138 | 0.9384 | 2.1689 | 1.7027 |
| 127 | aC-CONH(CH$_2$)$_2$N | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 128 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | 4.6466 | 0.6155 | 4.1474 | 0.4567 | 4.4159 | 1.6027 | 3.7676 | 3.0591 | 4.6311 | 0.6682 | 4.6749 | 0.5460 | 4.6466 |
| 129 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | 4.4830 | **** | 1.8210 | **** | 1.7980 | **** | 1.4651 | **** | 1.3212 | **** | 1.4017 | **** | 4.4830 |
| 130 | aC-SO$_2$NH$_n$ ($n$>=0;$n$<3) | 4.9480 | **** | 4.1501 | **** | 4.8739 | **** | 3.8738 | **** | 4.8151 | **** | 4.9756 | **** | 4.9480 |

[a] The symbols $EUA_{C\ 2j}$, $EUA_{NC\ 2j}$, $ERA_{C\ 2j}$, $ERA_{NC\ 2j}$, $EFW_{C\ 2j}$, $EFW_{NC\ 2j}$, $ESW_{C\ 2j}$, $ESW_{NC\ 2j}$, $ENS_{C\ 2j}$, $ENS_{NC\ 2j}$, $EAS_{C\ 2j}$, and $EAS_{NC\ 2j}$ represent the contributions ($D_j$) of the second-order groups for the corresponding properties.

**Table S7. MG Method Based Property Models Analysed Using Simultaneous Regression Method: Third-Order Groups and their Contributions [a] for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), and PCO**

| | Group | $LC50(FM)_{3k}$ | $LC50(DM)_{3k}$ | $LD50_{3k}$ | $LogWs_{3k}$ | $BCF_{3k}$ | $PEL_{3k}$ | $PCO_{3k}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | HOOC-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | 0.9173 | **** | -0.4207 | 0.2040 | **** | -0.9038 | -0.3757 |
| 2 | $NH_n$-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | **** | **** | **** | 1.0560 | **** | **** | **** |
| 3 | $NH_2$-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | -0.2468 | **** | **** | **** | **** |
| 4 | OH-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | 0.2831 | 1.3300 | **** | **** | -0.1802 |
| 5 | OH-$(CH_p)_k$-O-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 6 | OH-$(CH_p)_k$-S-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 7 | OH-$(CH_p)_k$-$NH_x$-$(CH_n)_m$-OH ($m,k$>0; $p,n,x$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 8 | $CH_p$-O-$(CH_n)_m$-OH ($m$>2; $n,p$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 9 | $NH_2$-$(CH_n)_m$-$NH_2$ ($m$>2; $n$ in 0..2)) | -0.8275 | **** | 0.3173 | **** | **** | -0.9748 | **** |
| 10 | $NH_k$-$(CH_n)_m$-$NH_2$ ($m$>2; $k$ in 0..1; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 11 | SH-$(CH_n)_m$-SH ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 12 | NC-$(CH_n)_m$-CN ($m$>2) | -1.3554 | **** | 1.3010 | 1.3500 | **** | -1.1729 | **** |
| 13 | COO-$(CH_n)_m$-OOC ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 14 | aC-$(CH_n$=$CH_m)_{cyc}$ (fused rings) ($n,m$ in 0..1) | 0.1769 | 0.1711 | 0.2622 | 0.5630 | 0.2127 | -0.6066 | 0.5918 |
| 15 | aC-aC (different rings) | 0.5827 | 2.5396 | 0.1456 | -0.6550 | 0.3719 | -0.0226 | **** |
| 16 | aC-$CH_{ncyc}$ (different rings) ($n$ in 0..1) | 0.1573 | -0.7346 | 0.0603 | 0.5110 | 0.2323 | **** | **** |
| 17 | aC-$CH_{ncyc}$ (fused rings) ($n$ in 0..1) | 0.1206 | 1.1726 | 0.1955 | 0.3110 | 0.0839 | 0.8204 | 0.0211 |
| 18 | aC-$(CH_n)_m$-aC (different rings) ($m$>1; $n$ in 0..2) | -0.4732 | **** | -0.4598 | -1.1200 | -3.6829 | **** | **** |
| 19 | aC-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | 1.5711 | 1.0355 | -0.0306 | -0.3286 | **** | **** | **** |
| 20 | $CH_{cyc}$-$CH_{cyc}$ (different rings) | **** | **** | -0.7228 | 0.7183 | **** | **** | 0.2372 |
| 21 | $CH_{cyc}$-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 22 | $CH_{multiring}$ | 0.1112 | -1.1600 | -0.2905 | 0.2640 | 0.7080 | 0.4625 | 0.0488 |
| 23 | $C_{multiring}$ | 0.4663 | -0.3169 | -0.0516 | 0.2870 | -0.5604 | -0.8159 | -0.0500 |
| 24 | aC-$CH_m$-aC (different rings) ($m$ in 0..2) | -0.2180 | -0.5207 | -0.1477 | 0.3680 | 0.3005 | -2.2183 | 0.0870 |
| 25 | aC-$(CH_m$=$CH_n)$-aC (different rings) ($m,n$ in 0..2) | **** | -1.4543 | -0.2549 | -1.1124 | 0.0332 | **** | **** |
| 26 | $(CH_m$=$C)_{cyc}$-CH=CH-$(C$=$CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** |
| 27 | $(CH_m$=$C)_{cyc}$-$CH_p$-$(C$=$CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** |
| 28 | aC-CO-aC (different rings) | -0.2216 | **** | 0.3273 | 0.9580 | -0.4462 | **** | **** |
| 29 | aC-$CH_m$-CO-aC (different rings) ($m$ in 0..2) | **** | **** | -0.3630 | -0.7354 | 0.5276 | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 30 | aC-CO-(C=CH$_n$)$_{cyc}$ (different rings) ($n$ in 0..1) | **** | **** | -0.0893 | -0.6605 | **** | **** | **** |
| 31 | aC-CO-CO-aC (different rings) | **** | **** | **** | -0.0946 | **** | **** | **** |
| 32 | aC-CO$_{cyc}$ (fused rings) | 0.6771 | 1.0063 | 0.0602 | 0.0311 | 0.2109 | -1.4914 | -0.2966 |
| 33 | aC-CO-(CH$_n$)$_m$-CO-aC (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** |
| 34 | aC-CO-CH$_{ncyc}$ (different rings) ($n$ in 0..1) | **** | **** | **** | -0.5784 | **** | **** | **** |
| 35 | aC-CO-NH$_n$-aC (different rings) ($n$ in 0..1) | 0.0254 | **** | -0.3798 | -0.8170 | -0.2883 | **** | **** |
| 36 | aC-NH$_n$CONH$_m$-aC (different rings) ($n$,$m$ in 0..1) | 2.2685 | 1.1010 | **** | 0.6880 | -0.7634 | **** | **** |
| 37 | aC-CO-N$_{cyc}$ (different rings) | **** | **** | 0.6053 | -0.8920 | **** | **** | **** |
| 38 | aC-S$_{cyc}$ (fused rings) | -1.3941 | 0.6642 | 0.1930 | 0.0047 | 0.2622 | -0.4700 | **** |
| 39 | aC-S-aC (different rings) | -2.3824 | **** | -0.6622 | 0.0690 | -2.1514 | **** | **** |
| 40 | aC-PO$_n$-aC (different rings) ($n$ in 0..4) | **** | **** | **** | **** | **** | **** | **** |
| 41 | aC-SO$_n$-aC (different rings) ($n$ in 1..4) | **** | 1.2103 | -0.0401 | 0.0305 | **** | **** | **** |
| 42 | aC-NH$_{ncyc}$ (fused rings) ($n$ in 0..1) | 0.5676 | 0.7420 | 0.1005 | -0.1850 | 0.3119 | 0.6063 | **** |
| 43 | aC-NH-aC (different rings) | 0.6705 | **** | -0.0227 | -0.4270 | 0.6307 | -0.7435 | **** |
| 44 | aC-(C=N)$_{cyc}$ (different rings) | **** | **** | -0.1243 | 0.5607 | -1.1970 | **** | **** |
| 45 | aC-(N=CH$_n$)$_{cyc}$ (fused rings) ($n$ in 0..1) | -0.3559 | -0.4921 | 0.2448 | -0.0985 | -0.3981 | **** | **** |
| 46 | aC-(CH$_n$=N)$_{cyc}$ (fused rings) ($n$ in 0..1) | **** | **** | -0.0772 | 0.3102 | **** | **** | **** |
| 47 | aC-O-CH$_n$-aC (different rings) ($n$ in 0..2) | **** | **** | -0.2144 | 1.0144 | **** | **** | **** |
| 48 | aC-O-aC (different rings) | 0.3988 | 1.2922 | -0.0525 | -0.7136 | -0.0772 | -0.0011 | **** |
| 49 | aC-CH$_n$-O-CH$_m$-aC (different rings) ($n$,$m$ in 0..2) | **** | **** | -0.8111 | 0.2160 | **** | **** | **** |
| 50 | aC-O$_{cyc}$ (fused rings) | 0.1768 | -0.3242 | 0.1886 | -0.6070 | 0.3585 | **** | **** |
| 51 | AROM.FUSED[2] | 0.3214 | 1.4255 | -0.1320 | -0.0759 | -0.1167 | 0.5784 | -0.0172 |
| 52 | AROM.FUSED[2]s$^1$ | 0.8251 | 1.4112 | 0.2442 | 0.0980 | 0.3046 | -0.5039 | -0.0424 |
| 53 | AROM.FUSED[2]s$^2$ | 1.0648 | 1.5552 | -0.1073 | -0.0711 | -0.0191 | 2.0554 | 0.0075 |
| 54 | AROM.FUSED[2]s$^2$s$^3$ | 0.5430 | 6.5271 | 0.3370 | 0.1059 | 0.2984 | **** | 0.1336 |
| 55 | AROM.FUSED[2]s$^1$s$^4$ | 0.0865 | **** | -0.2060 | -0.2716 | 0.1158 | **** | **** |
| 56 | AROM.FUSED[2]s$^1$s$^2$ | **** | **** | 0.1782 | -0.0567 | -1.1589 | **** | **** |
| 57 | AROM.FUSED[2]s$^1$s$^3$ | **** | **** | 0.4300 | -0.0732 | 0.3121 | **** | **** |
| 58 | AROM.FUSED[3] | 1.0422 | 0.7528 | -0.1893 | -0.1255 | -0.0916 | 0.7487 | **** |
| 59 | AROM.FUSED[4a] | **** | **** | -0.3160 | -0.6696 | -0.0827 | 0.0083 | **** |
| 60 | AROM.FUSED[4a]s$^1$ | **** | **** | -0.5015 | 0.2968 | 0.0928 | **** | **** |
| 61 | AROM.FUSED[4a]s$^1$s$^4$ | **** | **** | 1.1864 | 0.3025 | **** | **** | **** |
| 62 | AROM.FUSED[4p] | 2.2573 | 2.2734 | -0.2774 | 0.0500 | -0.2365 | 0.0083 | **** |
| 63 | AROM.FUSED[4p]s$^3$s$^4$ | **** | **** | 0.2710 | 2.1472 | **** | **** | **** |
| 64 | PYRIDINE.FUSED[2] | 0.2293 | 0.8970 | 0.0213 | 0.5540 | -1.5994 | **** | **** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 65 | PYRIDINE.FUSED[2-iso] | **** | 1.0799 | 0.6366 | 1.3326 | **** | **** | **** |
| 66 | PYRIDINE.FUSED[4] | 1.4995 | 2.5289 | 0.1623 | 2.0700 | -0.5698 | -1.0718 | **** |
| 67 | aC-N-$CH_{cyc}$ (different rings) | **** | **** | 0.3013 | -0.1996 | **** | **** | **** |
| 68 | N multiring | -0.0623 | 6.7397 | 0.0393 | 0.4788 | **** | **** | 0.1011 |
| 69 | $N_{cyc}$-$(CH_2)_3$-$N_{cyc}$ (different rings) | **** | **** | -0.1340 | -0.3802 | **** | **** | **** |
| 70 | aC-$COCH_2CH_2$-aC (different rings) | **** | **** | -0.2681 | -0.5406 | **** | **** | **** |
| 71 | aC-O-$(CH_2)_2$-$N_{cyc}$ (different rings) | **** | **** | -0.6543 | **** | **** | **** | **** |
| 72 | aC-CH(OH)$(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** |
| 73 | Ncyc-$(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | -0.3193 | **** | **** | **** | **** |
| 74 | aC-$CONHCH_2$-$CH_{cyc}$ (different rings) | **** | **** | 0.4260 | **** | **** | **** | **** |

[a] The symbols $LC50(FM)_{3k}$, $LC50(DM)_{3k}$, $LD50_{3k}$, $LogWs_{3k}$, $BCF_{3k}$, $PEL_{3k}$, $PCO_{3k}$ represent the contributions ($E_k$) of the third-order groups for the corresponding properties. Note that there are no third-order contributions for GWP, ODP and AP since these properties belong to chemicals whose structure is defined by first-order groups alone.

**Table S7 (continued). MG Method Based Property Models Analysed Using <u>Simultaneous Regression Method</u>: Third-Order Groups and their Contributions [a] for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| | Group | $EUA_{C\ 3k}$ | $EUA_{NC3k}$ | $ERA_{C\ 3k}$ | $ERA_{NC3k}$ | $EFW_{C\ 3k}$ | $EFW_{NC\ 3k}$ | $ESW_{C\ 2j}$ | $ESW_{NC\ 3k}$ | $ENS_{C\ 3k}$ | $ENS_{NC\ 3k}$ | $EAS_{C\ 3k}$ | $EAS_{NC\ 3k}$ | $EUA_{C\ 3k}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HOOC-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 2 | $NH_n$-$(CH_n)_m$-COOH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 3 | $NH_2$-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 4 | OH-$(CH_n)_m$-OH ($m$>2, $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 5 | OH-$(CH_p)_k$-O-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 6 | OH-$(CH_p)_k$-S-$(CH_n)_m$-OH ($m,k$>0; $p,n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 7 | OH-$(CH_p)_k$-$NH_x$-$(CH_n)_m$-OH ($m,k$>0; $p,n,x$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 8 | $CH_p$-O-$(CH_n)_m$-OH ($m$>2; $n,p$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 9 | $NH_2$-$(CH_n)_m$-$NH_2$ ($m$>2; $n$ in 0..2)) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 10 | $NH_k$-$(CH_n)_m$-$NH_2$ ($m$>2; $k$ in 0..1; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 11 | SH-$(CH_n)_m$-SH ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 12 | NC-$(CH_n)_m$-CN ($m$>2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 13 | COO-$(CH_n)_m$-OOC ($m$>2; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 14 | aC-$(CH_n{=}CH_m)_{cyc}$ (fused rings) ($n,m$ in 0..1) | 1.1453 | **** | 0.3743 | **** | 0.7155 | **** | 1.1082 | **** | 1.1576 | **** | 1.1447 | **** | 1.1453 |
| 15 | aC-aC (different rings) | -0.3745 | 0.3502 | 0.0337 | -0.6127 | -0.2123 | -0.5070 | -0.2114 | -0.5153 | -0.1751 | -0.1023 | 0.2783 | 0.4820 | -0.3745 |
| 16 | aC-$CH_{ncyc}$ (different rings) ($n$ in 0..1) | -1.6763 | 1.2288 | -0.5860 | 0.2367 | -0.3896 | 1.8325 | -0.5974 | 2.1670 | -0.3506 | -0.1667 | -1.0244 | 2.0064 | -1.6763 |
| 17 | aC-$CH_{ncyc}$ (fused rings) ($n$ in 0..1) | 0.5989 | 0.2239 | 0.4792 | 0.1614 | 0.7203 | -0.0600 | 0.8116 | -0.3012 | 0.4990 | 0.1098 | 0.4442 | 0.4050 | 0.5989 |
| 18 | aC-$(CH_n)_m$-aC (different rings) ($m$>1; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 19 | aC-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 20 | $CH_{cyc}$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 21 | $CH_{cyc}$-$(CH_n)_m$-$CH_{cyc}$ (different rings) ($m$>0; $n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 22 | $CH_{multiring}$ | 0.2918 | 1.4253 | 0.5641 | 0.1846 | -0.2737 | 1.5142 | -0.8597 | 3.2326 | 0.3179 | 0.5498 | -0.2432 | 2.5758 | 0.2918 |
| 23 | $C_{multiring}$ | -2.3245 | -1.6313 | -0.8597 | -3.5534 | -0.4698 | -1.5278 | -0.8958 | -2.7675 | 0.8845 | -1.0496 | -0.0873 | 1.9041 | -2.3245 |
| 24 | aC-$CH_m$-aC (different rings) ($m$ in 0..2) | -0.2959 | 0.9218 | 0.3732 | 0.4699 | 0.3818 | 1.1639 | 0.0487 | 1.1364 | 1.0739 | 1.2047 | 0.9533 | 1.3399 | -0.2959 |
| 25 | aC-$(CH_m{=}CH_n)$-aC (different rings) ($m,n$ in 0..2) | -1.1581 | **** | -3.5109 | **** | -1.3118 | **** | -2.4897 | **** | -0.6754 | **** | -0.7573 | **** | -1.1581 |
| 26 | $(CH_m{=}C)_{cyc}$-CH=CH-$(C{=}CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 27 | $(CH_m{=}C)_{cyc}$-$CH_p$-$(C{=}CH_n)_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 28 | aC-CO-aC (different rings) | 2.9341 | **** | **** | **** | **** | **** | 2.3524 | **** | 2.2595 | **** | **** | **** | 2.9341 |
| 29 | aC-$CH_m$-CO-aC (different rings) ($m$ in 0..2) | **** | **** | 1.8412 | **** | 1.6105 | **** | **** | **** | **** | **** | 2.1609 | **** | **** |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | aC-CO-(C=CH$_n$)$_{cyc}$ (different rings) ($n$ in 0..1) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 31 | aC-CO-CO-aC (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 32 | aC-CO$_{cyc}$ (fused rings) | 0.5575 | 0.2980 | 0.4436 | -0.2414 | 0.5009 | -0.0452 | 0.1402 | 0.2220 | 0.4805 | 0.2377 | 0.4643 | 0.4622 | 0.5575 |
| 33 | aC-CO-(CH$_n$)$_m$-CO-aC (different rings) ($m$>0; $n$ in | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 34 | aC-CO-CH$_{ncyc}$ (different rings) ($n$ in 0..1) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 35 | aC-CO-NH$_n$-aC (different rings) ($n$ in 0..1) | **** | **** | **** | **** | **** | **** | **** | 0.9066 | **** | **** | **** | **** | **** |
| 36 | aC-NH$_n$CONH$_m$-aC (different rings) ($n$,$m$ in 0..1) | 0.8217 | **** | **** | **** | **** | **** | 1.5104 | **** | **** | **** | **** | **** | 0.8217 |
| 37 | aC-CO-N$_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 38 | aC-S$_{cyc}$ (fused rings) | 0.6815 | 0.6537 | 1.0950 | 0.3350 | 0.8548 | 1.2864 | 0.3803 | -0.0805 | 2.1013 | -0.4347 | 1.9715 | -0.0939 | 0.6815 |
| 39 | aC-S-aC (different rings) | -1.5222 | **** | **** | **** | -1.4878 | **** | **** | **** | -0.8046 | **** | **** | **** | -1.5222 |
| 40 | aC-PO$_n$-aC (different rings) ($n$ in 0..4) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 41 | aC-SO$_n$-aC (different rings) ($n$ in 1..4) | 1.2240 | **** | 1.8319 | **** | 2.2857 | **** | 1.8029 | **** | 2.3910 | **** | 2.0459 | **** | 1.2240 |
| 42 | aC-NH$_{ncyc}$ (fused rings) ($n$ in 0..1) | 1.4988 | 0.4368 | 1.3033 | 0.4905 | 0.9940 | 0.3361 | 1.5455 | -0.6428 | 1.6993 | 0.6135 | 1.4767 | 0.6416 | 1.4988 |
| 43 | aC-NH-aC (different rings) | 4.2975 | 0.6868 | 3.7434 | 1.0236 | 3.7193 | 0.5016 | 3.0226 | 0.2959 | 3.5508 | 0.4208 | 3.2482 | 0.6133 | 4.2975 |
| 44 | aC-(C=N)$_{cyc}$ (different rings) | 1.0499 | 0.4529 | **** | -1.7816 | -2.6381 | 0.6892 | **** | -1.3698 | -4.0912 | -0.3454 | **** | 0.3300 | 1.0499 |
| 45 | aC-(N=CH$_n$)$_{cyc}$ (fused rings) ($n$ in 0..1) | -0.8892 | 0.9142 | -0.4406 | 0.9818 | 0.2535 | -0.2085 | 1.6274 | 1.4884 | -0.0745 | -0.0211 | 0.1517 | -0.1760 | -0.8892 |
| 46 | aC-(CH$_n$=N)$_{cyc}$ (fused rings) ($n$ in 0..1) | **** | **** | -0.6373 | **** | **** | **** | 1.5519 | **** | **** | **** | -2.1856 | **** | **** |
| 47 | aC-O-CH$_n$-aC (different rings) ($n$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 48 | aC-O-aC (different rings) | 0.2963 | 0.0230 | 0.0970 | 0.5422 | 0.0342 | -0.0853 | -0.1136 | -0.8360 | 0.2707 | 1.6559 | 0.6668 | 1.6485 | 0.2963 |
| 49 | aC-CH$_n$-O-CH$_m$-aC (different rings) ($n$,$m$ in 0..2) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 50 | aC-O$_{cyc}$ (fused rings) | 0.9721 | 0.8889 | 1.2643 | -0.6400 | 0.3761 | 0.1663 | 0.7579 | 0.5447 | 0.9779 | 0.6311 | 0.7324 | -1.8948 | 0.9721 |
| 51 | AROM.FUSED[2] | -0.0316 | -0.5677 | 0.5201 | -0.6876 | 0.1126 | -1.1200 | 0.1693 | -0.6070 | -0.0125 | -0.7338 | 0.0912 | -4.2858 | -0.0316 |
| 52 | AROM.FUSED[2]s$^1$ | 0.0476 | 0.8444 | -0.0543 | 1.2897 | -0.2536 | 0.9398 | 0.2944 | 0.6790 | 0.0195 | 0.6832 | 0.0904 | 1.2549 | 0.0476 |
| 53 | AROM.FUSED[2]s$^2$ | 0.3853 | -0.1511 | -0.4057 | 0.3161 | -0.1155 | -0.3720 | 0.1988 | -0.4414 | 0.3824 | -0.4358 | 0.2054 | -0.7488 | 0.3853 |
| 54 | AROM.FUSED[2]s$^2$s$^3$ | -3.2511 | -1.5652 | 0.1564 | 1.6397 | -0.6531 | -1.2387 | -2.7482 | -0.9316 | -2.5448 | 0.0525 | -2.2312 | -0.0208 | -3.2511 |
| 55 | AROM.FUSED[2]s$^1$s$^4$ | 0.0712 | **** | 0.4869 | **** | 0.1806 | **** | 0.7572 | **** | 0.3580 | **** | 0.2650 | **** | 0.0712 |
| 56 | AROM.FUSED[2]s$^1$s$^2$ | 0.6821 | -3.9117 | -0.5138 | -2.6314 | -0.2806 | -2.8689 | 1.3343 | -5.4893 | 1.0399 | -0.2754 | 0.7327 | -4.8866 | 0.6821 |
| 57 | AROM.FUSED[2]s$^1$s$^3$ | -1.3732 | **** | -1.1393 | **** | -0.6221 | **** | -0.8382 | **** | -1.5324 | **** | -1.8774 | **** | -1.3732 |
| 58 | AROM.FUSED[3] | -0.2936 | -0.5401 | 0.5174 | -0.5452 | 0.1298 | -1.2250 | -0.1360 | -1.0354 | -0.0145 | -0.7079 | 0.5637 | -4.3457 | -0.2936 |
| 59 | AROM.FUSED[4a] | 0.1157 | **** | 0.4646 | 1.0427 | 0.1454 | 0.5616 | 0.8427 | 0.1918 | -0.4812 | 0.0101 | 0.2781 | 0.0172 | 0.1157 |
| 60 | AROM.FUSED[4a]s$^1$ | 1.0468 | **** | 1.8126 | **** | 1.6412 | **** | 1.9925 | **** | 1.0138 | **** | 1.4593 | **** | 1.0468 |
| 61 | AROM.FUSED[4a]s$^1$s$^4$ | -0.7641 | **** | -2.4896 | **** | -2.0882 | **** | -1.8129 | **** | -2.0344 | **** | -1.8514 | **** | -0.7641 |
| 62 | AROM.FUSED[4p] | -0.5768 | -0.4101 | -0.4202 | -0.6491 | -0.0859 | -0.5568 | 0.2636 | -0.4714 | -0.5330 | -0.4819 | -0.6143 | -2.2533 | -0.5768 |
| 63 | AROM.FUSED[4p]s$^3$s$^4$ | 3.7567 | **** | 0.1636 | **** | -6.4758 | **** | 7.5539 | **** | -4.8061 | **** | -2.2659 | **** | 3.7567 |
| 64 | PYRIDINE.FUSED[2] | -0.7884 | **** | -0.5325 | **** | -0.6751 | **** | 0.4602 | **** | -0.6388 | **** | -0.2095 | **** | -0.7884 |
| 65 | PYRIDINE.FUSED[2-iso] | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 66 | PYRIDINE.FUSED[4] | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 67 | aC-N-$CH_{cyc}$ (different rings) | 3.4700 | **** | 1.2220 | **** | -0.1785 | **** | -2.0455 | **** | **** | **** | 0.8371 | **** | 3.4700 |
| 68 | N multiring | 0.0626 | 0.0942 | 0.2739 | -0.9523 | -0.5110 | 0.8827 | -0.0914 | 0.4870 | -0.0101 | 1.4318 | 0.2564 | 1.5434 | 0.0626 |
| 69 | $N_{cyc}$-$(CH_2)_3$-$N_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 70 | aC-$COCH_2CH_2$-aC (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 71 | aC-O-$(CH_2)_2$-$N_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 72 | aC-$CH(OH)(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 73 | Ncyc-$(CH_2)_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |
| 74 | aC-$CONHCH_2$-$CH_{cyc}$ (different rings) | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** | **** |

[a] The symbols $EUA_{C\ 3k}$, $EUA_{NC\ 3k}$, $ERA_{C\ 3k}$, $ERA_{NC\ 3k}$, $EFW_{C\ 3k}$, $EFW_{NC\ 3k}$, $ESW_{C\ 3k}$, $ESW_{NC\ 3k}$, $ENS_{C\ 3k}$, $ENS_{NC\ 3k}$, $EAS_{C\ 3k}$, and $EAS_{NC\ 3k}$ represent the contributions ($E_k$) of the third-order groups for the corresponding properties.

**Table S8. CI Method Based Property Models: Atom Contributions and Model Constants for the Properties÷ $LC_{50}$(FM), $LC_{50}$(DM), $LD_{50}$, $LogW_s$, BCF, PEL(OSHA-TWA), PCO, GWP, ODP, and AP.**

| Parameter | $LC_{50}$ (FM) | $LC_{50}$ (DM) | $LD_{50}$ | $LogW_s$ | BCF | PEL | PCO | GWP | ODP | AP |
|---|---|---|---|---|---|---|---|---|---|---|
| a(H) | -0.14756 | -0.18656 | -0.01047 | -0.18973 | -0.01672 | -0.12954 | 0.079188 | 1.67752 | -3.37822 | 0.086429 |
| a(Cl) | 0.080051 | -0.11963 | 0.152121 | -12.3996 | 0.089494 | 0.428893 | 0.749693 | 3.423638 | -11.9804 | 0.23872 |
| a(Br) | 0.183264 | 0.044889 | 0.572806 | -27.3702 | -0.21653 | 0.550528 | 1.136445 | 4.522534 | -19.2469 | **** |
| a(F) | -0.06844 | 0.205501 | 0.197228 | -6.47571 | 0.117109 | -0.22263 | 0.620819 | 2.521654 | -5.02663 | 0.296694 |
| a(I) | 0.30059 | 2.003523 | 0.30088 | -42.9538 | -0.31876 | 1.033486 | **** | 1.781068 | **** | **** |
| a(N) | -0.01828 | -0.02395 | 0.139871 | -4.71295 | -0.35926 | 0.645083 | 0.09317 | **** | **** | 0.097458 |
| a(O) | -0.08007 | -0.21041 | 0.004962 | -5.13567 | -0.36447 | 0.224864 | 0.126631 | 0.424024 | -3.62951 | 0.044628 |
| a(P) | -0.20932 | 2.16253 | 1.181 | -10.462 | 0.723115 | -0.97773 | **** | **** | **** | **** |
| a(S) | 0.07225 | 0.247054 | 0.241384 | -11.058 | 0.054859 | 1.524772 | -0.24978 | **** | **** | **** |
| a(C) | 0.285658 | 0.246449 | 0.052218 | -4.42181 | 0.083797 | 0.325169 | -0.03181 | -3.05993 | -1.59188 | -0.10248 |
| a(Si) | **** | **** | -0.02767 | -11.0138 | 0.562317 | 1.052495 | **** | **** | **** | **** |
| *b* | 0.035109 | 0.253883 | 0.018103 | -0.09321 | 0.283525 | -0.42181 | -0.10486 | -0.01877 | 8.967731 | -0.01884 |
| *c* | 0.184584 | 0.024195 | -0.02677 | -0.10952 | -0.2022 | 0.241452 | 0.005087 | -1.52848 | 0.22623 | -0.12489 |
| *d* | 0 | 0 | 0 | 0 | 0.892444 | 2.998465 | -0.25708 | -0.52073 | 3.298083 | -0.27126 |

**Table S8 (continued). CI Method Based Property Models: Atom Contributions and Model Constants for the Properties÷ $EUA_C$, $EUA_{NC}$, $ERA_C$, $ERA_{NC}$, $EFW_C$, $EFW_{NC}$, $ESW_C$, $ESW_{NC}$, $ENS_C$, $ENS_{NC}$, $EAS_C$, and $EAS_{NC}$.**

| Parameter | $EUA_C$ | $EUA_{NC}$ | $ERA_C$ | $ERA_{NC}$ | $EFW_C$ | $EFW_{NC}$ | $ESW_C$ | $ESW_{NC}$ | $ENS_C$ | $ENS_{NC}$ | $EAS_C$ | $EAS_{NC}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a(H) | -0.0035 | -0.0056 | -0.0215 | 0.0073 | -0.0337 | -0.0046 | 0.0682 | 0.0429 | -0.0527 | -0.0333 | -0.0407 | -0.0142 |
| a(Cl) | -0.3184 | -0.3843 | -0.5551 | -0.5190 | -0.4765 | -0.3676 | -0.3006 | -0.5570 | -0.3129 | -0.3199 | -0.3358 | -0.2783 |
| a(Br) | -0.0954 | -0.2498 | -0.6338 | -0.3712 | -0.4577 | -0.0662 | 0.1593 | -0.1070 | -0.2447 | -0.0237 | -0.1894 | -0.0239 |
| a(F) | 0.0843 | -0.1457 | -0.1594 | -0.1494 | -0.0821 | -0.1904 | -0.1000 | -0.1074 | -0.0716 | -0.1681 | -0.0844 | -0.1567 |
| a(I) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| a(N) | -0.1865 | -0.0900 | -0.3080 | -0.2023 | -0.3263 | -0.0970 | 0.0363 | 0.0149 | -0.2198 | -0.1179 | -0.2186 | -0.0806 |
| a(O) | -0.0263 | -0.0875 | -0.1808 | -0.1550 | 0.0258 | 0.0332 | 0.2636 | 0.2549 | -0.0090 | -0.0565 | 0.0153 | -0.0367 |
| a(P) | 0.6655 | -0.8595 | -0.0865 | -0.7380 | -0.2442 | -1.5692 | 0.6766 | -1.7886 | 0.2774 | -1.0916 | 0.5614 | -0.9972 |
| a(S) | -0.1146 | -0.0669 | -0.3863 | 0.0192 | -0.4479 | -0.2211 | 0.1692 | -0.0612 | -0.2393 | -0.1541 | -0.2917 | -0.1470 |
| a(C) | -0.0373 | -0.0205 | -0.0748 | -0.0580 | -0.0651 | -0.0401 | -0.0824 | -0.1454 | 0.1092 | 0.1268 | 0.0478 | 0.0751 |
| a(Si) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| *b* | 0.0605 | 0.0293 | 0.1630 | 0.0838 | 0.0670 | -0.0763 | -0.2888 | -0.0771 | -0.1154 | -0.1819 | -0.0141 | -0.1722 |
| *c* | 0.0334 | -0.0235 | 0.0567 | -0.1104 | 0.1590 | 0.1340 | 0.2296 | -0.0066 | 0.2944 | 0.2490 | 0.0721 | 0.1618 |
| *d* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |