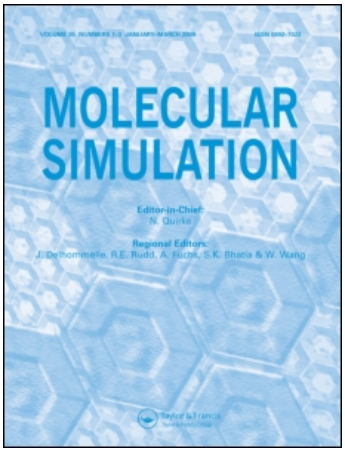

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Non-Linear QSAR Treatment of Genotoxicity

M. Karelson[a]; S. Sild[a]; U. Maran[a]
[a] Department of Chemistry, University of Tartu, Tartu, Estonia

To cite this Article Karelson, M. , Sild, S. and Maran, U.(2000) 'Non-Linear QSAR Treatment of Genotoxicity', Molecular Simulation, 24: 4, 229 — 242
To link to this Article: DOI: 10.1080/08927020008022373
URL: http://dx.doi.org/10.1080/08927020008022373

Molecular Simulation, 2000, Vol. 24, pp. 229–242
Reprints available directly from the publisher
Photocopying permitted by license only

NON-LINEAR QSAR TREATMENT OF GENOTOXICITY

M. KARELSON*, S. SILD and U. MARAN

Department of Chemistry, University of Tartu, Jakobi Street 2, Tartu, 51014 Estonia

(Received April 1999; accepted October 1999)

The nonlinear QSAR approach using the Chebyshev polynomial expansion and neural networks has been applied for the prediction of genotoxicity of compounds. The mutagenic toxicity of heteroaromatic and aromatic amines, measured by the Ames test, was correlated with the molecular descriptors calculated from the molecular structures using quantum-chemical methods. The quantitative models obtained were compared with the results of the linear QSAR treatment. The descriptors appearing in the models reveal the importance of mutagenic interactions of heteroaromatic amines *via* hydrogen bonding, of effects induced by the solvent, and of the size of compound. The dependence of molecular descriptors on environmental effects and on molecular conformations was analysed.

Keywords: Mutagenicity; Chebyshev polynomial expansion; neural networks; prediction of genotoxicity

INTRODUCTION

Quantitative structure-activity relationships (QSAR) have been used extensively to develop models for the prediction of toxicity by relating it to chemical structure [1–4]. The carcinogenicity and mutagenicity of compounds are important and closely related chronic toxicity measures [5–7]. The experimental assessment of carcinogenicity is often expensive and time-consuming. However, several tests allow easy detection of mutagenicity. One of the most widely used is a bacterial test based on the *Salmonella typhimurium* strain, introduced by Ames *et al.* [8,9]. In the present communication, structure-activity relationships were developed for a data

* Corresponding author.

set of experimental mutagenic potencies of aromatic amines towards *Salmonella typhimurium* TA98+S9 microsomial preparation (expressed by the logarithm of the number of revertants per nanomole, log R). This set has been previously compiled by Basak *et al.*, and involves 95 aromatic and heteroaromatic amines [10].

Traditional QSAR is based on the (multiple) linear relationships between the observed biological activity and one or several molecular descriptors. The latter may be derived from experimental data or from certain theoretical models, that relate the activity to the molecular structure. Different molecular descriptors (hydrophobic, electronic, steric, topological indices *etc.*) have been used to describe the QSAR in biological systems. In most treatments, only a limited set of descriptors has been used as the initial pool from which the best correlating scales are selected. Using larger variety of descriptors should enable the derivation of new models that describe more reliably the molecular activity. For instance, quantum-chemical calculations can be a major source of theoretical descriptors based on the structure and electronic properties of compounds [11]. Successful statistical QSPR models have been derived using very large pools of descriptors to predict various physical properties. Those include boiling point, melting point, flash point and gas chromatographic retention characteristics of organic compounds, glass transition temperatures of polymers, critical micelle concentrations of surfactants and other properties [12, 13]. This success encouraged us to use a large descriptor space as initial pool for the QSAR description of data on the Ames test of mutagenicity.

In most cases, the QSAR equations have been developed in the form of a linear equation where the coefficients for different descriptor scales are derived using multiple linear regression technique. However, the dependence of a property (biological activity) on the molecular descriptors may be intrinsically non-linear. This situation requires some non-linear approach in the development of the respective QSAR. In this communication, we consider two such approaches. First, the possibility to improve the multiple linear QSAR equations is explored using the Chebyshev polynomial expansion of the most significant descriptors. Secondly, the non-linear dependence is pursued implicitly using the back propagation neural networks.

It has been argued that the non-linear dependence between the observable property and molecular descriptors may arise due to the environmental (solvent) effects on the structure and properties of compounds and from the conformational changes in the molecules [14]. Therefore, an analysis of the solvent and conformational effects on various types of molecular descriptors was carried out.

METHODOLOGY

Altogether 618 molecular descriptors were generated for the 95 molecules listed in Table I, using the *CODESSA* software [15, 16]. These descriptors have been conventionally divided into several classes. *Constitutional*

TABLE I The comparison of the experimental data and neural network predictions of log R values

Structure	*log R (exp)*	*log R (calc)*	Δ^a
2-aminophenazine	0.550	0.790	0.240
2-amino-3′-nitrobiphenyl	−0.890	−0.868	0.022
2,6-dichloro-1,4-phenylenediamine	−0.690	−1.228	0.538
2-aminobiphenyl	−1.490	−0.621	0.869
4-amino-3′-nitrobiphenyl	1.020	0.521	0.499
3,3′-dimethylbenzidine	0.010	−0.148	0.158
2-methoxy-5-methylaniline	−2.050	−2.170	0.120
2,4-dinitroaniline	−2.000	−1.948	0.052
benzidine	−0.390	0.025	0.415
2-amino-7-nitrofluorene	3.000	2.810	0.190
2-chloroaniline	−3.000	−2.549	0.451
3,3′-dimethoxybenzidine	0.150	0.118	0.032
4,4′-diaminophenylsulfide	0.310	−0.050	0.360
4-amino-2′-nitrobiphenyl	−0.920	−0.452	0.468
1-aminocarbazole	−1.040	−1.042	0.002
1-aminophenanthrene	2.380	1.941	0.439
2-amino-7-acetamidofluorene	1.180	0.725	0.455
2-aminofluoranthene	3.230	2.852	0.378
1-aminopyrene	1.430	2.780	1.350
2-aminoanthracene	2.620	2.388	0.232
4,4′-methylenebis(*o*-fluoroaniline)	0.230	0.538	0.308
4-ethoxyaniline	−2.300	−2.374	0.074
2-amino-4-methylphenol	−2.100	−2.750	0.650
3-aminocarbazole	−0.480	−0.869	0.389
2-amino-5-nitrophenol	−2.520	−2.802	0.282
3-amino-2′-nitrobiphenyl	−1.300	−1.421	0.121
4-aminophenyldisulfide	−1.030	−0.787	0.243
1-aminofluoranthene	3.350	2.775	0.575
2-amino-4-chlorophenol	−3.000	−2.084	0.916
4,4′-methylenebis(*o*-ethylaniline)	−0.990	−1.287	0.297
9-aminoanthracene	0.870	1.636	0.766
3-amino-4′-nitrobiphenyl	0.690	0.351	0.339
4-chloro-2-nitroaniline	−2.220	−2.333	0.113
4,4′-methylenedianiline	−1.600	−1.519	0.081
3-aminophenanthrene	3.770	2.146	1.624
9-aminophenanthrene	2.980	1.860	1.120
3-amino-3′-nitrobiphenyl	−0.550	−0.655	0.105
7-aminofluoranthene	2.880	2.916	0.036
2,8-diaminophenazine	1.120	0.907	0.213
1,6-diaminophenazine	0.200	−0.006	0.206
2,4-diaminoisopropylbenzene	−3.000	−2.701	0.299
3-aminofluorene	1.930	2.257	0.327
5-aminoquinoline	−2.000	−2.097	0.097

TABLE I (Continued)

Structure	*log R (exp)*	*log R (calc)*	Δ[a]
6-aminochrysene	1.830	2.644	0.814
4-chloroaniline	−2.520	−2.339	0.181
1,7-diaminophenazine	0.750	0.951	0.201
2-amino-4′-nitrobiphenyl	−0.620	0.145	0.765
3-methoxy-4-methylaniline	−1.960	−2.237	0.277
1-aminoanthracene	1.180	2.163	0.983
3-amino-*a,a,a*-trifluorotoluene	−0.800	−1.432	0.632
4-aminocarbazole	−1.420	−1.228	0.192
2,4′-diaminobiphenyl	−0.920	−1.087	0.167
1,9-diaminophenazine	0.040	−0.125	0.165
4,4′-methylenebis(*o*-isopropylaniline)	−1.770	−1.381	0.389
8-aminoquinoline	−1.140	−2.162	1.022
4-aminopyrene	3.160	2.570	0.590
4-amino-biphenyl	−0.140	0.804	0.944
2,4-difluoroaniline	−2.700	−2.260	0.440
1-aminophenazine	−0.010	0.033	0.043
3-aminoquinoline	−3.140	−1.962	1.178
2,4-diamino-*n*-butylbenzene	−2.700	−2.789	0.089
6-aminoquinoline	−2.670	−1.967	0.703
3,3′-dichlorobenzidine	0.810	0.617	0.193
2,2′-diaminobiphenyl	1.520	0.989	0.531
2,4,5-trimethylaniline	−1.320	−2.010	0.690
2,5-dimethylaniline	−2.400	−2.427	0.027
4-aminophenylether	−1.140	−0.525	0.615
2-aminophenanthrene	2.460	2.384	0.076
4-bromoaniline	−2.700	−1.694	1.006
2,7-diaminophenazine	3.970	1.988	1.982
4-cyclohexylaniline	−1.240	−1.621	0.381
1-amino-4-nitronaphthalene	−1.770	−1.365	0.405
3-aminofluoranthene	3.310	3.017	0.293
2-bromo-7-aminofluorene	2.620	3.015	0.395
2-amino-1-nitronaphthalene	−1.170	−1.300	0.130
2-aminonaphthalene	−0.670	−1.127	0.457
3-aminofluorene	0.890	1.348	0.458
2-aminopyrene	3.500	2.710	0.790
4-aminofluorene	1.130	1.342	0.212
1-aminofluorene	0.430	1.258	0.828
8-aminofluoranthene	3.800	3.044	0.756
1-aminonaphthalene	−0.600	−1.283	0.683
3,4′-diaminobiphenyl	0.200	−0.669	0.869
4,4′-ethylenebis(aniline)	−2.150	−0.970	1.180
4-fluoroaniline	−3.320	−2.872	0.448
4-methoxy-2-methylaniline	−3.000	−2.045	0.955
4-phenoxyaniline	0.380	−0.491	0.871
2-bromo-4,6-dinitroaniline	−0.540	−1.449	0.909
2-hydroxy-7-aminofluorene	0.410	1.190	0.780
4-chloro-1,2-phenylenediamine	−0.490	−0.704	0.214
4-amino-4′-nitrobiphenyl	1.040	1.407	0.367
3,3′-diaminobiphenyl	−1.300	−0.341	0.959
2-aminocarbazole	0.600	−0.509	1.109
2,7-diaminofluorene	0.480	1.160	0.680
2,4-dimethylaniline	−2.220	−2.372	0.152

[a] $\Delta = |\log R(\text{calc}) - \log R(\text{exp})|$.

descriptors depend only on the chemical composition of the molecule and describe simple dependencies such as the additivity of molecular properties from constant fragment contributions. *Topological* descriptors are derived from the two-dimensional structural formula of the molecule. They reflect the molecular connectivity and branching of the molecule. *Electrostatic* descriptors quantify the details of the charge distribution in the molecules. In many cases, they are also related to the molecular topology and composition. In this work, the atomic partial charges were calculated using different empirical schemes that proceed from the electronegativities of atoms [16]. In addition, the Mulliken charges obtained from the quantum mechanical calculations were used. *Geometrical* descriptors reflect the three-dimensional structure and the shape of the molecule. A large number of molecular and local quantities characterising the reactivity, shape and binding of a molecule as well as its molecular fragments and substituents can be defined as *quantum chemical descriptors*. The information related to the electronic wave function and electron distribution can be efficiently used in deriving numerous descriptors. Those include the bond orders for specific types of bonds, chemical reactivity indices and polarizabilities of molecules. *Thermodynamic descriptors*, which include the heat of formation, entropy and heat capacity of the compound, can be derived together with other quantum-chemical information using the MOPAC package [17]. *Solvational* descriptors are also based on the quantum mechanical calculations and can be obtained using SCRF2.2 program of self-consistent reaction field model implemented in this package [18]. In this work, all calculations were carried out using the AM1 semiempirical parameterisation [19].

The derivation of the best multiple linear regression QSAR model has been carried out using the forward selection of descriptor scales as follows [16]. First, all descriptors were checked to ensure that numerical values are available for each structure and that there is a variation in these values. Descriptors for which the values were not available for every structure or having a constant value for all structures in the data set were discarded. Thereafter, the one-parameter correlation equations for each descriptor were calculated. In order to reduce further the number of descriptors in the set, several criteria were applied to eliminate other redundant descriptors. These included the Fischer criterion F for the one-parameter correlation with a given descriptor, the squared correlation coefficient for this one-parameter equation and the Student t-value for a given descriptor. Then the following proceedings were followed. (i) Starting from the top descriptor from the pre-selected list, the two-parameter correlations were calculated for all pairs of descriptors. (ii) The 400 best descriptor pairs showing the highest

F-values in the two-parameter correlations were selected and processed further as the working sets. (iii) Thereafter, each remaining descriptor, if not collinear with the descriptors already included, was added to the selected working set. (iv) After all descriptors had been applied one-by-one, the improvement of the correlations was examined in comparison with the best correlation of the previous rank. If the improvement was insignificant then the procedure was completed and the best equation obtained as the best correlation of the previous rank. Otherwise, the 400 best extended working sets, *i.e.*, the sets with the highest *F*-values, were submitted to the procedure from step (iii).

A similar treatment was carried out by extending the pool of descriptors with their nonlinear transformations obtained as the first five terms in the Chebyshev polynomial expansion. This expansion has been defined by the following generic formula

$$P_n(\bar{x}) = \frac{1}{2^n n!} \frac{d^n}{d\bar{x}^n} (\bar{x}^2 - 1)$$

where $\bar{x}$ are the centred and normalised descriptors defined as follows

$$\bar{x} = \frac{D - \bar{D}}{\sigma_D}$$

where D is the value of the natural descriptor for a given compound, $\bar{D}$ is the mean value and σ_D – the dispersion of this descriptor for a given set of compounds. At each step of the forward selection of scales, the Chebyshev polynomial terms of different order for the best scales from the previous step were applied as additional descriptors. The normalisation procedure ensures the approximate orthogonality of the modified descriptors, as the individual terms of Chebyshev polynomial are orthogonal in the interval $[-1, 1]$.

Finally, the neural network technique was applied for the development of a QSAR model. Three-layer neural network, made from six input units, five hidden units, and one output unit was chosen to represent nonlinear QSAR model ($6 \times 5 \times 1$) for $\log R$. The neural network was trained with the back propagation algorithm on a training set consisting of 50 structures [20]. In order to stop the training before overfitting, a validation set (consisting of 45 structures) was used. It was rather important to have both the testing and the validation sets as diverse as possible to ensure the stability of the resulting network. This selection was also essential to obtain the best predictions of the validation set data.

Initially, descriptors from six parameter linear models were used as input to the neural network. Notably, such input did not give substantial

improvement over the linear model. However, a noticeable improvement was achieved after optimisation of the neural network by replacing the descriptors from the linear model one by one at input nodes. The number of nodes was not changed. Since the training of neural networks involves significant amount of computation, the number of iterations was limited. For the development of the final network, only a certain number ($10 \cdots 20$) of models (with the least error) were investigated in further details.

RESULTS AND DISCUSSION

The best multiple linear model for the description of $\log R$ from the Ames test for 95 aromatic and heteroaromatic amines had been obtained as follows [21]

$$\begin{aligned}\log R = &-(0.16 \pm 2.70) + (0.85 \pm 0.15)N_R + (1.36 \pm 0.24)10^{-4}\gamma \\ &- (3.40 \pm 0.68)10^{-2}\mathrm{HASA}_1 + (1.71 \pm 0.35)10^{-2}\mathrm{HDSA} \\ &+ (0.90 \pm 0.18)E_{\mathrm{tot}}(\mathrm{C—C}) \\ &+ (0.90 \pm 0.18)E_{\mathrm{tot}}(\mathrm{C—N}) \\ R^2 = 0.8344 \quad R^2_{\mathrm{cv}} = 0.8056 \quad s^2 = 0.658 \quad F = 73.91\end{aligned}$$

where N_R denotes the number of rings in the compounds, γ is the γ-polarizability, HASA_1 and HDSA are the hydrogen bond acceptor and hydrogen bond donor surface areas [22, 23], respectively. The two last descriptors are the maximum total interaction energies for a C—C bond and for a C—N bond in the molecule. Those are defined as the sums of the respective quantum chemically calculated Coulomb and exchange interactions between the two atoms involved. These descriptors may be related to the formation of highly reactive radical centres in the aromatic systems that affect the reproductory system of the cell or DNA. All descriptors except the first were calculated using the quantum-chemical AM1 method [19].

The application of different terms of Chebyshev polynomials did not lead to better correlations. For instance, the use of the linear and the square term on the number of rings, N_R, gave the squared correlation coefficient $R^2 = 0.683$. The best linear two-parameter correlation with two different descriptors (N_R and γ-polarizability) was better, with $R^2 = 0.717$. A similar result was observed in the case of higher order correlations. The quadratic terms on different descriptors (*e.g.*, Hansch $\log P$ parameter) have been frequently used in QSAR correlations [4]. However, in our case the use of

square and higher order nonlinear terms on descriptors did not improve the QSAR equations.

On the other hand, the neural network treatment of log R data gave better result than the multiple linear regression model. The best neural network model for log R has $R^2 = 0.895$. The following six descriptors were involved in this model: the molecular moment of inertia (I_C), Kier shape index ${}^3\kappa$ [24], the number of hydrogen acceptor sites in the molecule, the maximum valence of a nitrogen atom, the partial negative surface area ($PNSA_1$), and γ-polarizability of the molecule. The predicted log R values are listed in Table I and the plot of experimental values *vs.* calculated values of log R is given in Figure 1.

The theoretical molecular descriptors, including the quantum-chemical descriptors have been usually obtained for the isolated molecules. However, the molecular wave function and the respective electronic and structural properties of molecules may be affected by the solvation. Thus, we have attempted to estimate the importance of such effects on the numerical values of different descriptors. A set of 406 compounds incorporating a wide variety of organic molecules was used for this analysis (*cf.* the list in the *Appendix*). The pair correlation coefficient was calculated for each

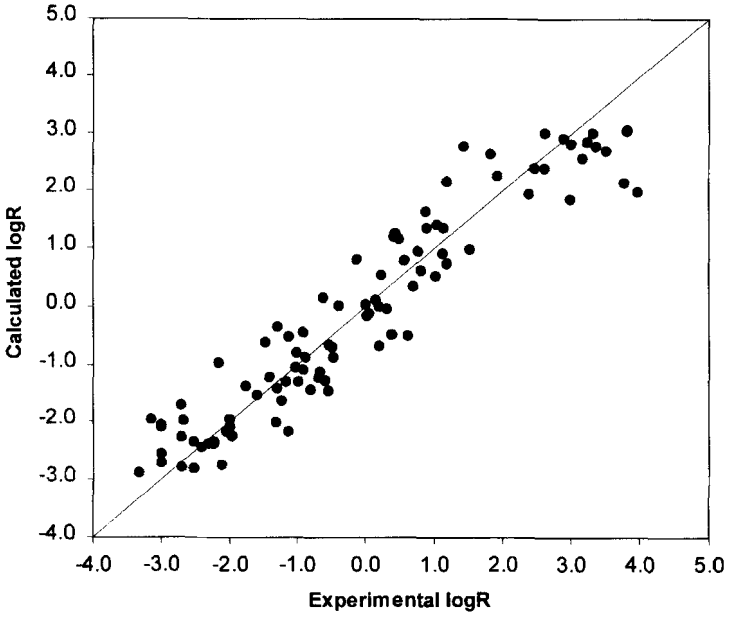

FIGURE 1 Experimental log R *vs.* calculated (neural network model) log R.

descriptor between the values belonging to the isolated molecule and to the molecule in a condensed medium. This medium was characterised within the self-consistent reaction field model by the dielectric constant of water ($\varepsilon = 80$) [25, 26]. In Table II, the distribution of different types descriptors is given according to this pair correlation coefficient. Interestingly, the correlation between the two types is very high ($R^2 > 0.99$) for most descriptors. Obviously, the perturbation by dielectric medium is similar in the case of different compounds and thus the linear relationships obtained with one set of descriptors will be preserved if another set is applied. This result is rather important as validating the use of quantum-chemical descriptors for isolated molecules in different condensed media.

The molecular descriptors can also depend on the conformation of the molecule. Whereas the constitutional and topological descriptors are insensitive to the conformational change, most of geometrical and charge-distribution related descriptors should be affected by these changes. In order to examine the conformational effects on the molecular descriptors, we carried out an analogous analysis as described above for the solvent-dependent descriptors. For a following series of compounds: $CH_3(CH_2)_nCH_3$, $CH_3(CH_2)_nOH$, $CH_3(CH_2)_nNH_2$, $CH_3(CH_2)_nCOOH$, $CH_3(CH_2)_nCN$, $HO(CH_2)_nOH$, $H_2N(CH_2)_nNH_2$, $HOOC(CH_2)_nCOOH$, $NC(CH_2)_nCN$, where $n = 2 \cdots 5$, two sets of molecular descriptors were calculated. The first of them corresponds to the full-*trans* conformation of the hydrocarbon skeleton of compounds, whereas the other is calculated for the molecules in the folded conformational minimum. The data on the distribution of the pair correlation coefficients between these two sets of descriptors is given in Table III. It is interesting to note that the conformational change of molecule has much larger influence on the values of molecular descriptors than the change of the medium. Therefore, it has to be concluded that the choice of the "correct" conformation of the molecule from which the molecular descriptors are calculated is vital for the development of legitimate QSAR equations. W. G. Richards and co-workers have also pointed out the importance of statistical weighting over the

TABLE II The correlation between the molecular descriptors calculated for the isolated molecules and for the molecules in a dielectric medium with $\varepsilon = 80$, respectively

Descriptors	*N*	$R^2 < 0.95$	$R^2 < 0.99$
Geometrical	12	1	4
Quantum-chemical energy	162	1	4
Quantum-chemical charge	176	9	35
Molecular orbital related	64	0	5
Total	414	11	48

TABLE III The correlation between the molecular descriptors calculated for the molecules in full-*trans* and in the folded optimum conformation, respectively

Descriptors	*N*	$R^2 < 0.5$	$R^2 < 0.95$	$R^2 < 0.99$
Geometrical	15	4	7	0
Quantum-chemical Energy	120	9	27	86
Quantum-chemical charge	141	3	19	47
Molecular orbital related	64	11	37	20
Total	330	27	81	153

conformations at given temperature, using the respective Boltzmann factors [27, 28].

CONCLUSIONS

The QSAR treatment of mutagenic potencies of aromatic amines towards Salmonella typhimurium TA98+S (Ames test) indicates the importance of the charge distribution and bond energy related interactions on this toxic activity. The best 6-parameter linear model was substantially improved by implicit account for nonlinear effects using the back propagation neural networks. A more traditional approach using the higher exponential orders of descriptors *via* the respective Chebyshev polynomials did not lead to better results.

The analysis of the solvent and conformational dependence of different molecular descriptors revealed the importance of using correct conformation in the development of descriptors. The environmental (solvent) effects on descriptors are less important.

Acknowledgement

This work has been partially supported by the Estonian Science Foundation grants No. 3051 and No. 3572.

APPENDIX

Alphabetical list of compounds used in the analysis of solvent effects on the molecular descriptors.

Hydrocarbons: 1,2,3-Trimethylbenzene, 1,2,4-Trimethylbenzene, 1,3,5-Cycloheptatriene, 1,3,5-Trimethylbenzene, 1,3-Methylnaphthalene, 1,4-Methylnaphthalene, 1-Butene, 1-Butyne, 1-Ethylnaphthalene, 1-Heptene,

1-Heptyne, 1-Hexene, 1-Hexyne, 1-Methylcyclohexene, 1-Methylnaphthalene, 1-Nonene, 1-Octene, 1-Octyne, 1-Pentene, 1-Pentyne, 2,2,4-Trimethylpentane, 2,2,5-Trimethylhexane, 2,2-Dimethylbutane, 2,2-Dimethylpentane, 2,2-Dimethylpropane, 2,3,4-Trimethylpentane, 2,3-Dimethyl-1,3-butadiene, 2,3-Dimethylbutane, 2,3-Dimethylnaphthalene, 2,3-Dimethylpentane, 2,4-Dimethylpentane, 2,6-Dimethylnaphthalene, 2-Ethyltoluene, 2-Methyl-1,3-butadiene, 2-Methyl-1-pentene, 2-Methyl-2-butane, 2-Methylbutane, 2-Methylhexane, 2-Methylpentane, 2-Methylpropane, 2-Pentene, 3,3-Dimethylpentane, 3-Methyl-1-butene, 3-Methylheptane, 3-Methylhexane, 3-Methylpentane, 4-Ethyltoluene, 4-Isopropyltoluene, Acenaphthene fluorene, Benzene, Biphenyl, *cis*-1,2-Dimethylcyclohexane, Cyclohexane, Cyclohexene, Cyclopentane, Cyclopentene, Cyclopropane, Ethane, Ethene, Ethylbenzene, Indane, Isobutylbenzene, Isopropylbenzene, Methane, Methylcyclohexane, Methylcyclopentane, *m*-Xylene, Naphthalene, *n*-Butane, *n*-Butylbenzene, *n*-Decane, *n*-Heptane, *n*-Hexane, *n*-Hexylbenzene, *n*-Nonane, *n*-Octane, *n*-Pentane, *n*-Pentylbenzene, *n*-Propylcyclopentane, *o*-Xylene, Propane, Propene, Propylbenzene, Propyne, *p*-Xylene, *sec*-Butyl benzene, Styrene, *tert*-Butylbenzene, Toluene, *trans*-1,4-Dimethylcyclohexane, α-Methylstyrene.

Halogen derivatives: 1,1,1,2-Tetrachloroethane, 1,1,1-Trichloroethane, 1,1,2,2-Tetrachloroethane, 1,1,2-Trichloroethane, 1,1-Dichloroethane, 1,1-Dichloroethene, 1,2,3,4-Tetrachlorobenzene, 1,2,3,5-Tetrachlorobenzene, 1,2,3-Trichlorobenzene, 1,2,3-Trimethylbenzene, 1,2,4,5-Tetrachlorobenzene, 1,2,4-Trichlorobenzene, 1,2,4-Trimethylbenzene, 1,2-Dichlorobenzene, 1,2-Dichloroethane, 1,2-Dichloropropane, 1,3,5-Trichlorobenzene, 1,3-Dichlorobenzene, 1,3-Dichloropropane, 1,4-Dichlorobenzene, 1-Bromo-2-methylpropane, 1-Bromobutane, 1-Bromoheptane, 1-Bromohexane, 1-Bromooctane, 1-Bromopentane, 1-Bromopropane, 1-Chloro-2-propene, 1-Chlorobutane, 1-Chloroheptane, 1-Chlorohexane, 1-Chloropentane, 1-Chloropropane. 1-Cyanobutane, 1-Cyanopropane, 1-Ethylnaphthalene, 1-Heptanol, 1-Heptene, 1-Heptyne, 1-Hexanol, 1-Hexene, 1-Hexyne, 1-Iodobutane, 1-Iodohexane, 1-Iodopentane, 1-Iodopropane, 1-Methylcyclohexene, 1-Methylnaphthalene, 1-Naphthol, 1-Naphthylamine, 1-Nitrobutane, 1-Nitropentane, 1-Nitropropane, 1-Nonanol, 1-Nonene, 1-Octanol, 1-Octene, 1-Octyne, 1-Pentanol, 1-Pentene, 1-Pentyne, 1-Propanol, 2,2,2-Trifluoroethanol, 2,2,4-Trimethylpentane, 2,2,5-Trimethylhexane, 2,2-Dimethylbutane, 2,2-Dimethylpentane, 2,2-Dimethylpropane, 2,3,4-Trimethylpentanne, 2,3-Dimethyl-1,3-butadiene, 2,3-Dimethylbutane, 2,3-Dimethylnaphthalene, 2,3-Dimethylpentane, 2,3-Dimethylphenol, 2,3-Dimethylpyridine, 2,4-Dimethylpentane, 2,4-Dimethylphenol, 2,4-Dimethylpyridine,

2,5-Dimethylphenol, 2,5-Dimethylpyridine, 2,5-Dimethytetrahydrofuran, 2,6-Dimethylaniline, 2,6-Dimethylnaphthalene, 2,6-Dimethylphenol, 2,6-Dimethylpyridine, 2-Bromo-2-methylpropane, 2-Bromopropane, 2-Butanol, 2-Butenal, 2-Butoxyethanol, 2-Chloro-2-methylpropane, 2-Chloroaniline, 2-Chlorobutane, 2-Chlorophenol, 2-Chloropropane, 2-Chloropyridine, 2-Chlorotoluene, 2-Decanone, 2-Ethoxyethanol, 2-Ethylpyridine, 2-Ethyltoluene, 2-Flororphenol, 2-Heptanone, 2-Hexanone, 2-Iodophenol, 2-Methoxyaniline, 2-Methoxyethanol, 2-Methoxyphenol, 2-Methyl-1,3-butadiene, 2-Methyl-1-butanol, 2-Methyl-1-pentene, 2-Methyl-1-propanol, 2-Methyl-2-butanol, 2-Methyl-2-butene, 2-Methyl-2-pentanol, 2-Methyl-2-propanol, 2-Methyl-3-pentanol, 2-Methylbutane, 2-Methylhexane, 2-Methylpentane, 2-Methylpropane, 2-Methylpyrazine, 2-Methylpyridine, 2-Methylthiophene, 2-Methytetrahydrofuran, 2-Naphthol, 2-Naphthylamine, 2-Nitroaniline, 2-Nitrophenol, 2-Nitropropane, 2-Nitrotoluene, 2-Nonanone, 2-Octanone, 2-Pentanol, 2-pentanone, 2-Pentene, 2-Phenylethanol, 2-Propanol, 2-Propenol, 2-Undecanone, 3,3-Dimethylpentane, 3,4-Dimethylphenol, 3,4-Dimethylpyridine, 3,5-Dimethylphenol, 3,5-Dimethylpyridine, 3-Acetylpyridine, 3-Chloroaniline, 3-Chlorophenol, 3-Chloropyridine, 3-Cyanophenol, 3-Cyanopyridine, 3-Ethylphenol, 3-Ethylpyridine, 3-Formalpyridine, 3-Hexanol, 3-Methoxyaniline, 3-Methoxyphenol, 3-Methyl-1-butanol, 3-Methyl-1-butene, 3-methyl-2-butanone, 3-Methylbutanoic acid, 3-Methylheptane, 3-Methylhexane, 3-Methylpentane, 3-Methylpyridine, 3-Nitroaniline, 3-Nitrophenol, 3-Nitrotoluene, 3-Pentanol, 3-pentanone, 3-Phenyl-1-propanol, 4-Acetylpyridine, 4-Bromophenol, 4-Bromotoluene, 4-Chloro-3-methylphenol, 4-Chloroaniline, 4-Chlorophenol, 4-Cyanophenol, 4-Cyanopyridine, 4-Ethylphenol, 4-Ethylpyridine, 4-Ethyltoluene, 4-Fluorophenol, 4-Heptanone, 4-Hydroxybenzaldehyde, 4-Isopropyltoluene, 4-Methoxyaniline, 4-Methyl-2-pentanol, 4-Methyl-2-pentanone, 4-Methylacetophenone, 4-Methylbenzaldehyde, 4-Methylpyridine, 4-Nitroaniline, 4-Nitrophenol, 4-*n*-Propylphenol, 4-*tert*-Butylphenol, 5-Nonanone, Acenaphthene, Acetaldehyde, Acetic acid, Acetonitrile, Acetophenone, alfa-Methylstyrene, Ammonia, Benzaldehyde, Benzamide, Benzene, Benzonitrile, Benzotrifluoride, Benzyl alcohol, Biphenyl, Bromobenzene, Bromoethane, Bromomethane, Butanoic-acid, Butanone, Butyl acetate, Butylamine, Butyraldehyde, Chlorobenzene, Chloroethane, Chloromethane, *cis*-1,2-dichloroethylene, *cis*-1,2-dimethylcyclohexane, Cycloheptanol, Cyclohexane, Cyclohexanone, Cyclohexene, Cyclohexylamine, Cyclopentane, Cyclopentanol, Cyclopentanone, Cyclopentene, Cyclopropane, Decyl alcohol, Dibromomethane, Dichloromethane, Diethyl disulfide, Diethyl ether, Diethyl sulfide, Diethylamine, Diisopropyl ether, Diisopropyl sulfide, Diisopropylamine, Dimethylamine, Di-*n*-butyl ether, Di-*n*-butyl-

amine, Di-*n*-propyl ether, Di-*n*-propyl sulfide, Di-*n*-Propylamine, Ethane, Ethanethiol, Ethanol, Ethene, Ethyl acetate, Ethyl benzoate, Ethyl butanoate, Ethyl formate, Ethyl hexanoate, Ethyl pentanoate, Ethyl phenyl ether, Ethyl propanoate, Ethylamine, Ethylbenzene, Fluorene, Fluorobenzene, Formaldehyde, Halothane, Heptanal, Hexanal, Hexanoic acid, Hexyl acetate, Hexylamine, Indane, Iodobenzene, Iodoethane, Iodomethane, Isoamyl acetate, Isoamyl formate, Isobutyl acetate, Isobutyl formate, Isobutyl isotuanoate, Isobutylbenzene, Isobutyraldehyde, Isoflurane, Isopropyl acetate, Isopropyl formate, Isopropylbenzene, Methane, Methanol, Methoxyflurane, Methyl acetate, Methyl benzoate, Methyl butanoate, Methyl formate, Methyl hexanoate, Methyl pentanoate, Methyl phenyl ether, Methyl propanoate, Methylamine, Methylcyclohexane, Methylcyclopentane, *m*-Xylene, *N*,*N*-dimethylformamide, Naphthalene, *n*-Butane, *n*-butanethiol, *n*-Butylbenzene, *n*-Decane, *n*-Heptane, *n*-Heptylamine, *n*-Hexane, *n*-Hexylbenzene, Nitrobenzene, Nitroethane, Nitromethane, *N*-methylaniline, *N*-methylpipyridine, *n*-Nonane, *n*-Octane, Nonanal, *n*-Pentane, *n*-Pentyl propanoate, *n*-Pentylamine, *n*-Pentylbenzene, *n*-propanethiol, *n*-Propyl butanoate, *n*-Propyl propanoate, *n*-Propylcyclopentane, *o*-Cresol, Octanal, Octylamine, *o*-Toluidine, *o*-Xylene, *p*-Cresol, Pentachloroethane, Pentanal, Pentanoic acid, Phenol, Phenyl acetate, Phenyl methyl sulfide, Propane, Propanoic acid, Propanone, Propanonitrile, Propene, Propionaldehyde, Propyl acetate, Propyl formate, Propylamine, Propylbenzene, Propyne, *p*-Toluidine, *p*-Xylene, Pyridine, Quinoline, *sec*-Butyl benzene, Styrene, Teflurane, *tert*-Butylbenzene, Tetrachloroethylene, Tetracloromethane, Tetrafluoromethane, Tetrahydrofuran, Tetrahydropyran, Thiophene, Thiophenol, Toluene, *trans*-1,4-Dimethylcyclohexane, *trans*-2-Hexenal, Tribromomethane, Trichloroethylene, Tricloromethane, Triethylamine, Trimethylamine, Water.

References

[1] Goldberg, L. (Ed.), "Structure-Activity Correlations as a Predictive Tool in Toxicology", Hemisphere, Washington, DC, 1983.
[2] Hadzi, D. and Jerman-Blazic, J. (Eds.), "QSAR in Drug Design and Toxicology", Elsevier, Amsterdam, 1987.
[3] Kaiser, K. L. E. (Ed.), "QSAR in Environmental Toxicology", Reidel, Dordrecht, 1987.
[4] Hansch, C. and Leo, A., "Exploring QSAR, Fundamentals and Applications in Chemistry and Biology", ACS, Washington, DC, 1995.
[5] Frierson, M. R., Klopman, G. and Rosenkrantz, H. S. (1986). "Structure-activity relationships among mutagens and carcinogens: A review", *Environ. Mutagen.*, **8**, 283.
[6] Benigni, R., Andreoli, C. and Giuliani, A. (1989). "Quantitative structure-activity relationships: principles, and applications to mutagenicity and carcinogenicity", *Mut. Res.*, **221**, 197.
[7] Benigni, R. and Giuliani, A. (1996). "Quantitative structure-activity relationship (QSAR) studies of mutagens and carcinogens", *Med. Res. Rev.*, **16**, 267.

[8] McCann, J., Choi, E., Yamasaki, E. and Ames, B. N. (1975). "Detection of carcinogens as mutagens in the Salmonella/microsome test: Assay of 300 chemicals", *Proc. Nat. Acad. Sci. USA*, **72**, 5135.
[9] Ames, B. N. (1979). "Identifying environmental chemicals causing mutations and cancer", *Science*, **204**, 587.
[10] Basak, S. C., Grunwald, G. D. and Niemi, G. J., "Use of graphic-theoretic and geometrical molecular descriptors in structure-activity relationships", In: "*From Chemical Topology to Three-Dimensional Geometry*", Balaban, A. T. (Ed.), Plenum Press, New York, 1997, pp. 73–116.
[11] Karelson, M., Lobanov, V. S. and Katritzky, A. R. (1996). "Quantum-chemical descriptors in QSAR/QSPR studies", *Chem. Rev.*, **96**, 1027.
[12] Katritzky, A. R., Lobanov, V. S. and Karelson, M. (1995). "QSPR: The correlation and quantitative prediction of chemical and physical properties from structure", *Chem. Soc. Rev.*, **24**, 279.
[13] Katritzky, A. R., Karelson, M. and Lobanov, V. S. (1997). "QSPR as a means of predicting and understanding chemical and physical properties in terms of structure", *Pure and Appl. Chem.*, **69**, 245.
[14] Karelson, M. (1996). "Use of Semiempirical Quantum-Chemical Molecular Descriptors in QSAR/QSPR", *Abstr. Pap. Am. Chem. Soc.*, **211**, 154-COMP.
[15] CODESSA, Semichem, 7128 Summit, Shawnee, KS 66216.
[16] Katritzky, A. R., Lobanov, V. S. and Karelson, M., "CODESSA: Reference Manual, version 2.0.", *Gainesville, FL*, 1994.
[17] Stewart, J. J. P. "*MOPAC Program Package*", QCPE 1989, No. 455.
[18] Karelson, M., Tamm, T., Katritzky, A. R., Cato, S. J. and Zerner, M. C. (1989). "Application of Self-Consistent Reaction Field Method in Semiempirical Quantum-Chemical Calculations", *Tetrahedron Comp. Methodol.*, **2**, 295.
[19] Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P. (1985). "AM1: A new general purpose quantum mechanical molecular model", *J. Am. Chem. Soc.*, **107**, 3902.
[20] Mitchell, T. M., "Machine Learning", McGraw-Hill, New York, 1997.
[21] Maran, U., Katritzky, A. R. and Karelson, M. (1999). "A Comprehensive QSAR Treatment of the Genotoxicity of Heteroaromatic and Aromatic Amines", *Quant. Struct-Act. Relat.*, **18**, 617.
[22] Stanton, D. T. and Jurs, P. C. (1990). "Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Realtionship Studies", *Anal. Chem.*, **62**, 2323.
[23] Stanton, D. T., Egolf, L. M., Jurs, P. C. and Hicks, M. G. (1992). "Computer-Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles", *J. Chem. Inf. Comput. Sci.*, **32**, 306.
[24] Kier, L. B., "Computational Aspects of Molecular Connectivity and its Role in Structure-Property Modeling", In: "*Computational Chemical Graph Theory*", Rouvray, D. H. (Ed.), Nova Science Publishers, New York, 1990.
[25] Karelson, M. (1997). "Quantum Chemical Treatment of Molecules in Condensed Disordered Media", *Adv. Quant. Chem.*, **28**, 142.
[26] Karelson, M., "Molecular Properties and Spectra in Solution", In: "*Problem Solving in Computational Molecular Science: Molecules in Different Environments*", Wilson, S. and Diercksen, G. H. F. (Eds.), Kluwer Academic Publ., Dordrecht, 1997, pp. 353–387.
[27] Burt, C. and Richards, W. G. (1990). "Molecular similarity: The introduction of flexible fitting", *J. Comput.-Aided Mol. Design*, **4**, 231.
[28] Good, A. C. and Richards, W. G., "Explicit Calculation of 3D Molecular Similarity", In: "*3D QSAR in Drug Design*", Kubinyi, H., Folkers, G. and Martin, Y. C. (Eds.), Kluwer/Escom, Dordrecht, 1998, **2**, 321–338.