Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry
Bioorganic & Medicinal Chemistry 13 (2005) 6450–6463

QSAR modeling of blood:air and tissue:air partition coefficients using theoretical descriptors

Alan R. Katritzky,[a,*] Minati Kuanar,[a] Dan C. Fara,[a] Mati Karelson,[b] William E. Acree, Jr.,[c] Vitaly P. Solov'ev[d] and Alexandre Varnek,[e,*]

[a]*Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, FL 32611, USA*
[b]*Department of Chemistry, Tallinn University of Technology, Akadeemia tee 15, Tallinn 12618, Estonia*
[c]*Department of Chemistry, University of North Texas, Denton, TX 76203-5070, USA*
[d]*Institute of Physical Chemistry, Russian Academy of Sciences, Leninskii prosp. 31, 119991 Moscow, Russia*
[e]*Faculté de Chimie, 4 rue B. Pascal, Strasbourg 67000, France*

Received 27 April 2005; revised 29 June 2005; accepted 30 June 2005
Available online 3 October 2005

Abstract—Human blood:air, human and rat tissue (fat, brain, liver, muscle, and kidney):air partition coefficients of a diverse set of organic compounds were correlated and predicted using structural descriptors by employing CODESSA-PRO and ISIDA programs. Four and five descriptor regression models developed using CODESSA-PRO were validated on three different test sets. Overall, these models have reasonable values of correlation coefficients (R^2) and leave-one-out correlation coefficients (R^2_{cv}): $R^2 = 0.881$–0.983; $R^2_{cv} = 0.826$–0.962. Calculations with ISIDA resulted in models based on atom/bond sequences involving two to three atoms with statistical parameters that were similar to those of models obtained with CODESSA-PRO ($R^2 = 0.911$–0.974; $R^2_{cv} = 0.831$–0.936). A mixed pool of molecular and fragment descriptors did not lead to significant improvement of the models.

1. Introduction

Tuning the biosolubility of chemicals is a promising tool that has been extensively applied in recent drug discoveries and in anesthesiological and toxicological studies. In practice, the biological effects of inhaled chemicals have been related to exposure concentration (concentration in ambient air) but more fundamentally, the biological effects depend on chemical concentration in the target organ. On account of importance of the solubility of solutes in biological systems, several attempts have been made to correlate and to predict such solubilities, especially huge gaseous solutes in blood.[1,2]

Such solubilities are best explained in terms of partition coefficients, that is, ratios of concentrations achieved between two different media at equilibrium. Thus, solubility of a chemical in blood and tissue, indicated as blood:air and tissue:air partition coefficients, are important physicochemical properties for understanding the pharmacokinetics of organic solvents.

Partitioning of a compound to a specific tissue depends on the affinities of the specific compound for blood and for the tissue. The partitioning behavior of volatile organic compounds (VOCs) can be studied by the determination of partition coefficients in an in vitro closed vial system. Sato and Nakajima[3] used a vial-equilibration method, for the measurement of the partition coefficients of several aromatic hydrocarbons and ketones in water, blood, and oil. This experimental method needs no direct measurement of the concentration either in the liquid or in the air phase: it utilizes the ratio of gas chromatographic peak heights of air in the sample and reference vessels.

A modification by Gargas et al.[4] of the vial-equilibration technique of Sato and Nakajima[3] is currently being used by investigators for determining the blood:air and tissue:air partition coefficients for volatile solvents in tissue homogenates. These techniques are attractive because no elaborate experimental apparatus is required and the partition coefficient can be determined by measuring

Keywords: QSAR; CODESSA-PRO; ISIDA; Blood:air; Tissue:air; Partition coefficients; Structural descriptors.
* Corresponding authors. Tel.: +1 352 392 0554; fax: +1 352 392 9199 (A.R.K.); tel.: +33 390 241549; fax: 33 390 241545 (A.V.); e-mail addresses: katritzky@chem.ufl.edu; varnek@chimie.u-strasbg.fr

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.06.066

headspace concentrations in test vials and appropriate reference vials, with no requirement for measuring chemical concentrations in the test medium.[4] The partition coefficients from saline and olive oil to blood, liver, muscle, and fat tissues of rats were determined[4] for 55 organic compounds using the modified vial equilibration method. These authors also determined the human blood:air partition coefficients of 36 diverse organic compounds using the same method.[4]

Blood:air and tissue:air or tissue:blood partition coefficients are important indicators of pulmonary uptake and distribution of volatile chemicals in mammalian systems.[5–8] Such coefficients constitute an integral component in the development of physiologically based pharmacokinetic (PBPK) models.[9–12]

Hetrick et al.[13] considered that the most sensitive parameters in a physiological model associated with the tissue concentration of chemicals are the blood:air or tissue:air partition coefficients. Subsequently, detailed PBPK models have been developed to describe the metabolic rate constants, gas pharmacokinetics, and biotransformation of chemicals, and to correlate and predict toxicokinetics of chemicals in rat and human tissues.[14–35]

Payne and Kenny[36] reviewed several published models for calculating blood:air, tissue:air, or tissue:blood partition coefficients of volatile organic chemicals in human or rat tissues, from functions of their octanol–water partition coefficients or solubilities in vegetable oil and water. They compared the predicted human blood and tissue:air, rat blood, and tissue:air partition coefficients with experimental values for 12 chemicals, covering a wide range of lipophilicity. These authors used the ratio of predicted to experimental partition coefficients and their mean, R_{mean}, and the mean magnitude of the difference between predicted and experimental values of log P, E to assess the accuracy of each model. For the above set of 12 chemicals, the empirical equations of Meulenberg and Vijverberg[37] and the empirical solvation equation of Abraham and Weathersby[38] were the most accurate for human blood:air partition coefficients. It was found[36] that the prediction of rat blood:air partition coefficient was less accurate owing to difficulties in modeling the significant effects of protein binding. Overall, the empirical equation of Meulenberg and Vijverberg[37] was the best available for the prediction of rat blood:air partition coefficient in the above set of compounds.

To avoid experimental cost and time, several structure-based prediction models were proposed by various authors. In 1984, Fiserova-Bergerova et al.[5] reviewed the effects of biosolubility on pulmonary uptake and disposition of gases and vapors of lipophilic chemicals, reporting that the determination of tissue–gas (air) partition coefficients is tedious, and for highly soluble chemicals it brings along a large experimental error. It has been shown that molecular connectivity indexes provide reasonable information on tissue–gas partition coefficients of a series of aliphatic halogenated chemicals. A zero-order connectivity index is suggested for prediction of bioavailability of new chemicals with similar structures.[5] Several structure-based models have been used in investigating the relationship between chemical structure and biological toxicological, and environmental behavior of organic chemicals.[39–42]

Recently, structure-based models were reported by Basak et al.[43] for the prediction of rat tissue:air partition coefficients of 46 low molecular weight volatile chemicals. Structural descriptors (topostructural, topochemical, and 3-dimensional) and calculated octanol:water partition coefficient as descriptor variable and three linear regression methods (ridge regression (RR), principal component regression (PCR), and partial least-squares (PLS)) were used to develop QSPR models. The authors reported the prediction of human blood:air partition coefficients of 31 volatile organic chemicals[44] (R^2_{cv} = 0.874) and rat blood:air partition coefficients for a diverse set of 37 organic chemicals[45] (R^2_{cv} = 0.944) using the topochemical descriptors (TC). They concluded that for both QSPR predictions (rat tissue:air and human blood:-air partition coefficients) ridge regression (RR) is superior to PCR or PLS and topochemical descriptors are superior to other structural and experimental property data. A bilinear regression of rat tissue:air (R^2_{cv} = 0.894) and human blood:air (R^2_{cv} = 0.889) partition coefficients with oil:air and saline:air partition coefficients was developed,[43,44] as well as a trilinear equation employing four species-specific parameters (neutral lipid equivalent content of tissue, water equivalent content of tissue, protein content of tissue, and fraction of total protein involved in the partitioning process) and three chemical-specific parameters (vegetable oil:air, water:air, and protein:air partition coefficients).[46] In the latter approach, each additive term is a product of species-specific parameter(s) times a chemical-specific parameter.

QSPR models developed by our group have predicted aqueous solubility, vapor pressure, and water air partitioning coefficients of structurally diverse compounds.[47,48] Recently, we[49] have developed QSPR models for rat blood:air, saline:air, and olive oil:air partition coefficients using theoretical molecular descriptors. A training set of 100 diverse organic compounds were selected for the development of regression model for rat blood:-air partition coefficients. Subsequently, the saline:air and olive oil:air partition coefficients values were also modeled. The predictive power of these models tested on 33 compounds not included in the training set was satisfactory: correlation coefficients of 0.791 for rat blood:air, 0.794 for saline:air, and 0.846 for olive oil:air.

The present study aimed to build QSAR multiple regression models, to correlate and predict the human blood:-air, and human and rat tissue:air partition coefficients for a diverse set of 162 organic compounds. Two different QSAR approaches applied were based on: (i) molecular (CODESSA-PRO) and (ii) fragment (ISIDA) descriptors. The first approach uses the molecular descriptors calculated from the structure of a molecule either by quantum mechanical methods or by some empirical techniques. The second approach uses counts of substructural molecular fragments (SMF) as variables in a multiple regression analysis.

2. Methodology

2.1. CODESSA approach

CODESSA-PRO (comprehensive descriptors for structural and statistical analyses)[50] is a comprehensive program for developing QSAR/QSPR models. It includes diverse statistical structure–property–activity correlation techniques that can be used for the analysis in combination with calculated molecular descriptors. In particular, various algorithms based on stepwise statistical multilinear regression analysis are applicable for searching 'the best' multiparameter correlation in large spaces of natural molecular descriptors. All calculated descriptors used in the resulting multiparameter correlation equations are theoretical; hence, the value of the property/activity of interest can be predicted for unknown chemical structures. CODESSA-PRO methodology has shown promising results in medicinal chemistry applications.[51–53]

2.2. ISIDA approach

The method is based on the representation of molecular graph by a limited number of topological fragments and their contributions to a given property/activity. Computation of fragment descriptors does not require any knowledge of a 3D geometry and electronic structure of molecules. Regression coefficients related to structural fragments are more easily interpretable than those for some of the theoretical molecular descriptors. Molecular fragments are successfully used in the diversity analysis of large databases[54,55] and in structure–property studies.[56–61] The PASS method,[58,62] used for the estimation of a wide spectrum of biological activities, is also based on molecular fragments (augmented atoms). The disadvantage of QSPR methods based on fragments is related to the fact that they generally use more variables than those using 'traditional' descriptors, thus leading to smaller values of the Fischer criterion. However, our experience shows that the models based on fragment descriptors are at least as statistically stable and predictive as those molecular descriptors involved.[63–66] The success of the fragment approach in QSAR/QSPR studies depends on the diversity of structural fragments, as well as on the flexibility of atom/bond classification. The ISIDA program[67] represents a very flexible structure–property tool since it uses many different types of fragments (atom/bond sequences and augmented atoms) to build QSAR models involving linear and nonlinear fitting equations. Besides the QSAR module, ISIDA also includes the clustering and combinatorial modules which can be efficiently used for 'in silico' design of new compounds possessing desirable properties.

2.3. The present work

We report the QSAR modeling studies of the human blood:air, human tissue:air, and rat tissue:air partition coefficients for 162 diverse organic chemicals, performed with the ISIDA program, which uses the fragment descriptors, and with the CODESSA-PRO program, which applies up to 780 different constitutional, geometrical, topological, electrostatic, quantum chemical, and thermodynamic descriptors calculated for the whole molecule. All these descriptors are derived solely from chemical structure and do not require experimental data to be calculated. We have previously shown the advantage of using a combination of both the CODESSA-PRO and ISIDA approaches.[52]

3. Data sets and calculations

The experimental values for the following human and rat tissue:air and human blood:air partition coefficients of a diverse set 162 organic compounds were collected from several literature sources. Human blood:air partition coefficient data ($n = 139$) were collected from thirty literature sources.[3–5,34,68–93]

The human tissue partition coefficient data for fat:air ($n = 42$), brain:air ($n = 36$), liver:air ($n = 30$), muscle:air ($n = 39$), and kidney:air ($n = 34$) were selected from five studies.[31,38,39,90,94] Rat tissue:air partition coefficient data for fat:air ($n = 100$), brain:air ($n = 39$), liver:air ($n = 100$), muscle:air ($n = 97$), and kidney:air ($n = 27$) were collected from 30 articles.[14,15,17,20,21,26,29,31,34,38,75,95–113]

The average is taken when more than one experimental value is reported in the literature for specific human blood:air or human and rat tissue:air partition coefficients. The partition coefficient values were converted to their logarithm for the present study. The experimental partition coefficient values for blood:air and tissue:air in human and rats were given as supplementary material (see SM1).

3.1. CODESSA-PRO calculations

The chemical structures were drawn by using ISIS Draw 2.4[114] and pre-optimized using molecular mechanics force field method included in Hyperchem 7.0.[115] Final optimization was performed with MOPAC 7.0[116] (implemented in CODESSA-PRO) using the AM1 semi-empirical method. Molecular descriptors of different kinds such as constitutional, topological, geometrical, charge-related, semi-empirical, thermodynamic, molecular, atom, and bond type were calculated using the CODESSA-PRO program.[50] A gradient norm of 0.01 kcal/Å was applied to the geometry optimization.

The Best Multilinear Regression (BMLR) procedure available in the framework of the CODESSA-PRO was used to find the best correlation models from selected noncolinear descriptors. The BMLR selects the best two-parameter regression equation, the best three-parameter regression equation, etc., on the basis of the highest R^2 value in the stepwise regression procedure. During the BMLR procedure, the descriptor scales are normalized and centered automatically, and the final result is given in natural scales. The result obtained by BMLR is the best representation of the property in the given descriptors' pool.

Intercorrelations among the descriptors were checked by using BMLR options with maximum intercorrelation ($R^2 < 0.40$) for building the models throughout the entire studies. For the development of QSAR models, during the stepwise regression procedure, it is important to decide when to stop the addition of descriptors. An excessive number of descriptors leads to over-correlated equations that are difficult to interpret in terms of interactions and mechanisms. A simple procedure to control the proliferation of descriptor is the 'break point.' From analyses of the plot of the number of descriptors (k) involved versus the squared correlation coefficient (R^2), and the cross-validated square correlation coefficient (R^2_{cv}), using the values corresponding to the property models, it appears that the statistical improvement of the model is higher (steeper ascent of the relationship) until one point (the 'break point') and beyond that improvement is negligible. Consequently, the model corresponding to the break point shows the optimum number of descriptors to be used in modeling that property.

3.2. ISIDA calculations

The structure–property modeling was performed using the QSPR module of the ISIDA program, which realizes the SMF method.[67] The SMF method is based on the splitting of a molecular graph into fragments (subgraphs), and on the calculation of their contributions to a given property **Y**. Two different classes of fragments are used: 'sequences' (**I**) and 'augmented atoms' (**II**). Three subtypes **AB**, **A**, and **B** are defined for each class. For the fragments **I**, they represent sequences of atoms and bonds (**AB**), of atoms only (**A**), or of bonds only (**B**). Only the shortest paths from one atom to the other are used. For each type of sequence, the minimum (n_{min}) and maximum (n_{max}) number of constituted atoms must be defined. Thus, for partitioning **I**(**AB**, $n_{min} - n_{max}$), **I**(**A**, $n_{min} - n_{max}$), and **I**(**B**, $n_{min} - n_{max}$), the program generates 'intermediate' sequences involving n atoms ($n_{min} \leqslant n \leqslant n_{max}$). In the current version of ISIDA, $n_{min} \geqslant 2$ and $n_{max} \leqslant 6$. The number of types of sequences' of different lengths corresponding to $n_{min} = 2$ and $n_{max} = 6$ is equal to 15 for each of three sub-types **AB**, **A**, and **B**.

An 'augmented atom' represents a selected atom with its environment including either neighboring atoms and bonds (**AB**), or atoms only (**A**), or bonds only (**B**). Atomic hybridization (**Hy**) can be taken into account for augmented atoms of the **A**-type.

Once a molecular graph is split into constitutive fragments, any corresponding quantitative physical or chemical property **Y** is calculated from the fragment contributions using linear (FE1) or nonlinear (FE2) and (FE3) fitting equations:

$$\boldsymbol{Y} = a_o + \sum_i a_i N_i + \Gamma, \qquad \text{(FE1)}$$

$$\boldsymbol{Y} = a_o + \sum_i a_i N_i + \sum_i b_i (2N_i^2 - 1) + \Gamma, \qquad \text{(FE2)}$$

$$\boldsymbol{Y} = a_o + \sum_i a_i N_i + \sum_{i,k} b_{ik} N_i N_k + \Gamma, \qquad \text{(FE3)}$$

where a_i and $b_i(b_{ik})$ are fragment contributions and N_i is the number of fragments of i type. The a_o term is fragment independent. The a_i and $b_i(b_{ik})$ are the same for corresponding fragments for all compounds from the given set. An extra term $\Gamma = \sum c_m D_m$ can be used to describe any specific feature of the compound using external descriptors D_m (e.g., topological, electronic, etc.); by default $\Gamma = 0$. Eq. FE1 represents a molecular property as a linear combination of fragment contributions. Eq. FE2 representing the three first terms of Chebyshev polynomial[117] accounts for nonadditive effects related to individual fragments, whereas Eq. FE3 involves a cross-term N_i N_k, which accounts for the nonadditivity effects of two different fragments.

At the training stage, the program builds up to 147 structure–property models involving three linear and nonlinear fitting equations and 49 types of fragment descriptors (*batch* calculations). If some fragments are linearly dependent, they are treated as one extended fragment. Using the singular value decomposition (SVD) method,[118] the program fits the a_i and b_i terms in Eqs. FE1–FE3, calculates the corresponding statistical characteristics (correlation coefficient (R), standard deviation (s), Fischer's criterion (F), cross-validation correlation coefficient (R_{cv}), standard deviation of predictions (s_{cv}), Kubyni's criterion (FIT), R_H-factor of Hamilton, and matrix of pair correlations (covariation matrix) for the terms a_i and b_i), and performs statistical tests[119] to select the best models.

If some of the variables in Eqs. FE1–FE3, are linearly dependent or if a given fragment occurs in a relatively small number of molecules, the standard deviation $\Delta a_i(\Delta b_i)$ for the fragment contributions $a_i(b_i)$ can be large enough, leading to the corresponding t test ($t = a_i/\Delta a_i$) being smaller than the tabulated value (t_0). The following procedure is applied to improve the robustness of the models. First, the program selects the variable with the smallest $t < t_0$ and then it performs a new fitting, excluding that variable. This procedure is repeated until $t \geqslant t_0$ for all variables.

In general, ISIDA recognizes nine different types of bonds: single, double, triple (in cycle or in chain), aromatic bonds, and two types of coordination bonds. Therefore, the *EdChemS* editor of 2D structures (an element of the ISIDA system) has been used to normalize the bond types in the aromatic fragments originally presented as Kekule structures and to modify bond types in the cycles accounting for the difference between chains and cycles. Then, using the Edi*SDF editor* also included in ISIDA, structure data files (SDF) containing all 162 structures and experimental human and rat blood:air and tissue:air partition coefficients have been prepared and used further in structure property/activity modeling. All hydrogen atoms were omitted from ISIDA calculations.

4. Results and discussions

4.1. Analysis of experimental data

Analysis of the experimental data used with the DIVA program[120] shows that the studied human tissue:air partition coefficients are highly correlated with the corresponding rat tissue:air partition coefficients, with correlation coefficients of 0.96–0.99 (Fig. 1). This observation could help to establish empirical correlations that could complement the QSAR models in the prediction of the properties of compounds. In practice, this might be particularly useful to estimate human-tissue-related parameters from generally more easily available rat tissue parameters.

4.2. QSAR modeling

We have developed QSAR models based on two different approaches. Initially, two types of theoretical parameters that encode structural information were calculated for the whole data set of 162 compounds as follows: (i) 780 molecular descriptors using CODESSA-PRO software and (ii) 540 fragment descriptors by using the SMF method implemented in ISIDA program. Further, these parameters were used to build three pools of descriptors: (i) POOL1, which includes only molecular descriptors (780), (ii) POOL 2, which consists of 540 fragment descriptors, and (iii) POOL 3 by combining POOL 1 and POOL 2 (780 molecular and 58 fragment descriptors of the best models obtained by the SMF method). QSAR models were built for the blood:air and tissue:air partition coefficients using the three descriptor pools.

4.3. CODESSA calculations

The pool of 780 molecular descriptors was used to build the QSAR models of human and rat blood:air and tissue:air partition coefficients. For each training set QSAR models were developed using the BMLR procedure included in CODESSA-PRO software. To find a model with optimum number of descriptors, we built multiple parameter correlations with up to a maximum of 10 descriptors for each property. By analyzing the plots of the number of descriptors (k) versus the correlation coefficients (R^2) and the cross-validated correlation coefficient (R^2_{cv}) corresponding to those models, we selected the best/optimum descriptor models. The plots (see SM-2 Figs. 1a–k) show the break points, which indicate the optimum number of descriptors to be used in the corresponding QSAR models for POOL 1 descriptors. The best QSAR models (Eqs. 1–11) with 4 and 5 descriptors were thus obtained for the human and rat blood, fat, brain, liver, muscle, and kidney:air partition coefficients. The statistical characteristics of the regression model number of data points (n), number of descriptors (k), squared correlation coefficient (R^2), cross-validated squared correlation coefficient (R^2_{cv}), Fisher ratio (F), and squared standard deviation (s^2) together with descriptor coefficients are given in Table 1. The descriptors involved in the QSAR models are given in Table 2.

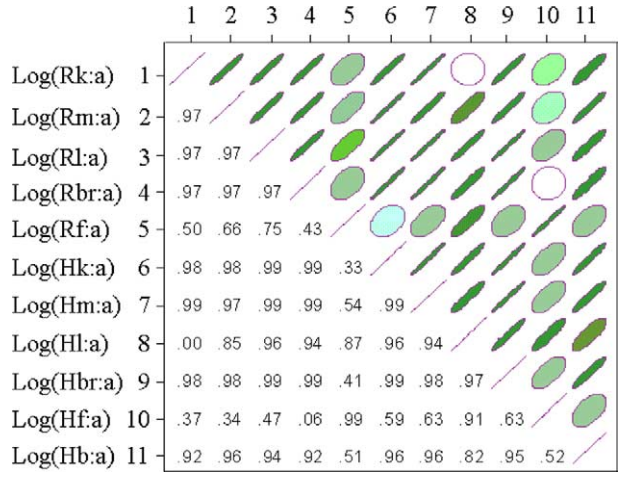


Figure 1. Correlation between experimental values of human-tissue and rat tissue/air partition coefficients used for structure–property modeling obtained with the Accelrys-DIVA program.

4.4. ISIDA calculations

For all the properties investigated, the best ISIDA models were selected from the *batch* calculations (see Section 2) based on the **I**(**AB** 2–3) fragmentation (shortest sequences of atom and bonds involving from two to three atoms) and linear Eq. FE1. Although the number of fragment descriptors involved in these models is larger than that in the models obtained with CODESSA-PRO, the number of data points per fragment descriptor varies from 3 to 6, which is still acceptable (Table 1). The models (Eqs. 12–22) have reasonable statistical parameters: $R^2 = 0.911$–0.974, $s^2 = 0.020$–0.10 $F = 28.3$–104.0, and $R^2_{cv} = 0.831$–0.936, which are similar to the values obtained for the CODESSA-PRO models. Figure 2 shows that property values calculated for the training set correspond well to the experimental data. The fragment descriptors involved in the models, as well as their coefficients in the fitting Eq. FE1, are given in Figure 3 and in Supplementary Material (SM3–SM6).

4.5. Combined approach

QSPR modeling of blood:air and tissue:air partition coefficients in humans and rats was performed using the fragment descriptors and molecular descriptors together (POOL 3 descriptors.) The large pool of 838 descriptors was used for QSPR modeling. The BMLR procedure was used to select the best set of descriptors.

The QSAR models for Log(*Hb*:*a*), Log(*Hf*:*a*), Log(*Hbr*:*a*), Log(*Hm*:*a*), Log(*Hk*:*a*), Log(*Rbr*:*a*), Log(*Rl*:*a*), Log(*Rm*:*a*), and Log(*Rk*:*a*) are identical to those obtained using solely the POOL 1 descriptors (Eqs. 1–3, 5, 6, 8–11, respectively). The QSAR model for Log(*Hl*:*a*) obtained using the POOL 3 descriptors is slightly improved in the correlation coefficient ($R^2 = 0.897$).

Table 1. Statistical parameters of QSAR models for human blood:air and human and rat tissue:air partition coefficients using CODESSA-PRO and ISIDA approaches

Eq. no.	Modeling property	QSAR Model: Equation	n	k	R^2	R^2_{cv}	F	s^2
POOL 1: CODESSA-PRO molecular descriptors								
1	Log(*Hb*:*a*)	Log(*Hb*:*a*) = (−1.654 ± 0.142) − (0.591 ± 0.027)D_2 + (0.004 ± 0.0002)D_{11} + (5.846 ± 0.322)D_{16} + (88.81 ± 6.446)D_{35} − (0.079 ± 0.008)D_1	139	5	0.907	0.897	259.6	0.140
2	Log(*Hf*:*a*)	Log(*Hf*:*a*) = (−0.263 ± 0.137) + (0.045 ± 0.003)D_{32} + (337.88 ± 25.37)D_{22} + (0.494 ± 0.042)D_3 + (0.511 ± 0.087)D_7	42	4	0.918	0.887	104.1	0.043
3	Log(*Hbr*:*a*)	Log(*Hbr*:*a*) = (5.740 ± 0.541) + (283.10 ± 11.12)D_{23} + (0.525 ± 0.045)D_{25} + (0.448 ± 0.043)D_3 + (0.753 ± 0.099)D_{15}	36	4	0.959	0.943	181.9	0.043
4	Log(*Hl*:*a*)	Log(*Hl*:*a*) = −(1.866 ± 0.340) + (0.056 ± 0.005)D_{24} + (0.692 ± 0.066)D_8 − (1.575 ± 0.159)D_{29} − (1.236 ± 0.187)D_{30}	34	4	0.881	0.826	46.4	0.043
5	Log(*Hm*:*a*)	Log(*Hm*:*a*) = (2.476 ± 0.377) + (626.04 ± 34.64) D_{22} − (0.218 ± 0.028)D_{28} + (0.483 ± 0.064)D_8 − (1.336 ± 0.307)D_5	39	4	0.916	0.897	92.90	0.092
6	Log(*Hk*:*a*)	Log(*Hk*:*a*) = (0.826 ± 0.154) + (571.95 ± 25.309)D_{17}+(18.56 ± 1.826)D_{18} − (0.418 ± 0.046)D_{26} + (1.689 ± 0.250)D_{35}	34	4	0.953	0.940	146.9	0.050
7	Log(*Rf*:*a*)	Log(*Rf*:*a*) = (0.930 ± 0.095) + (0.718 ± 0.040)D_8 + (0.201 ± 0.015)D_{20} − (0.367 ± 0.034)D_2 − (0.186 ± 0.020)D_{27} − (0.015 ± 0.002)D_{19}	99	5	0.916	0.904	203.6	0.062
8	Log(*Rbr*:*a*)	Log(*Rbr*:*a*) = −(1.637 ± 0.311) + (436.98 ± 28.27)D_{22} − (0.692 ± 0.061)D_2 + (0.005 ± 0.0005)D_{11} + (85.74 ± 14.48)D_{14} − (0.561 ± 0.143)D_{15}	59	5	0.905	0.879	101.3	0.060
9	Log(*Rl*:*a*)	Log(*Rl*:*a*) = −(1.351 ± 0.149) + (484.613 ± 22.747)D_{22} + (0.852 ± 0.042)$_{33}$ − (0.746 ± 0.046)D_2 + (162.41 ± 14.586)D_{40} − (0.494 ± 0.077)D_{41}	100	5	0.906	0.895	180.9	0.092
10	Log(*Rm*:*a*)	Log(*Rm*:*a*) = −(1.908 ± 0.167) + (5.548 ± 0.276)D_{16} + (0.005 ± 0.0002)D_{11} − (0.581 ± 0.036)D_2 + (75.527 ± 6.529)D_{14} − (0.013 ± 0.002)D_9	97	5	0.925	0.912	225.7	0.082
11	Log(*Rk*:*a*)	Log(*Rk*:*a*) = −(2.626 ± 0.252) + (4.117 ± 0.244)D_6 − (0.634 ± 0.055)D_{10} + (7.304 ± 0.689)D_{12} + (1.582 ± 0.242)D_{13}	27	4	0.983	0.962	325.5	0.011
POOL 2: ISIDA Fragment descriptors[a]								
12	Log(*Hb*:*a*)	—[c]	138[b]	24	0.938	0.909	74.6	0.10
13	Log(*Hf*:*a*)	—[d]	42	18	0.974	0.880	53.4	0.020
14	Log(*Hbr*:*a*)	—[d]	35[b]	12	0.960	0.918	50.0	0.048
15	Log(*Hl*:*a*)	—[d]	30	9	0.915	0.831	28.3	0.036
16	Log(*Hm*:*a*)	—[d]	38[b]	8	0.911	0.874	43.7	0.096
17	Log(*Hk*:*a*)	—[d]	34	8	0.924	0.894	45.1	0.090
18	Log(*Rf*:*a*)	—[c]	99	24	0.970	0.911	104.0	0.029
19	Log(*Rbr*:*a*)	—[d]	59	15	0.938	0.859	47.5	0.048
20	Log(*Rl*:*a*)	—[c]	100	21	0.939	0.891	60.4	0.073
21	Log(*Rm*:*a*)	—[c]	97	24	0.961	0.936	79.0	0.053
22	Log(*Rk*:*a*)	—[d]	27	6	0.946	0.914	73.3	0.040
POOL 3: CODESSA-PRO molecular and fragment descriptors[e]								
23	Log(*Rf*:*a*)	Log(*Rf*:*a*) = (0.054 ± 0.088) + (48.57 ± 3.767)D_{21} − (0.409 ± 0.039)D_{38} + (0.003 ± 0.0004)D_{11} − (0.106 ± 0.017)D_{39} + (0.547 ± 0.088)D_8	99	5	0.917	0.905	204.6	0.062

[a] The ISIDA models were based on the **I**(**AB** 2–3) fragmentation and linear Eq. FE1.
[b] Methane was excluded from the training set.
[c] Fragment descriptors and their contributions are given in Figure 3.
[d] Fragment descriptors and their contributions are given as supplementary data.
[e] The models for (Log(*Hb*:*a*), Log(*Hf*:*a*), Log(*Hbr*:*a*), Log(*Hm*:*a*), Log(*Hk*:*a*), Log(*Rbr*:*a*), Log(*Rl*:*a*), Log(*Rm*:*a*), and Log(*Rk*:*a*) are identical to those obtained with POOL 1 descriptors (Eqs. 1–3, 5, 6, 8–11, respectively).

Table 2. List of descriptors used (Eqs. 1–11 and 23) in the QSAR modeling of human blood:air and human and rat tissue:air partition coefficients using CODESSA-Pro and ISIDA approaches

Name of the descriptor	Symbol
Number of H atoms	D_1
Number of F atoms	D_2
Number of Cl atoms	D_3
Relative number of O atoms	D_4
Relative number of F atoms	D_5
Relative number of single bonds	D_6
Number of rings	D_7
Kier and Hall index (order 1)	D_8
Complementary information content (order 1)	D_9
Balaban index	D_{10}
Gravitation index (all bonds)	D_{11}
Polarity parameter (Zefirov)	D_{12}
Topographic electronic index (all bonds)	D_{13}
FPSA3 Fractional PPSA (PPSA-3/TMSA) (Zefirov PC)	D_{14}
WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (Zefirov PC)	D_{15}
HA dependent HDSA-2/SQRT(TMSA) (Zefirov PC)	D_{16}
HACA-2/TMSA (Zefirov PC)	D_{17}
FNSA-3 Fractional PNSA (PNSA-3/TMSA) (MOPAC PC)	D_{18}
WNSA-2 Weighted PNSA (PNSA2*TMSA/1000) (MOPAC PC)	D_{19}
HA dependent HDCA-1 (MOPAC PC)	D_{20}
HA dependent HDCA-1/TMSA (MOPAC PC)	D_{21}
HA dependent HDCA-2/TMSA (MOPAC PC)	D_{22}
HACA-2/TMSA (MOPAC PC)	D_{23}
Final heat of formation/# atoms	D_{24}
HOMO − 1 energy	D_{25}
LUMO energy	D_{26}
LUMO + 1 energy	D_{27}
HOMO–LUMO energy gap	D_{28}
Max atomic orbital electronic population	D_{29}
Min (>0.1) bond order for atom C	D_{30}
Vib enthalpy (300 K)	D_{31}
ALFA polarizability (DIP)	D_{32}
Randic index (order 1)	D_{33}
Partially charged surface area for atom H	D_{34}
Square root of charged surface area for atom H	D_{35}
Positively charged part of partial charged surface area (MOPAC PC)	D_{36}
HACA-2/TMSA (MOPAC PC) (all)	D_{37}
C—F	D_{38}
C—C—C	D_{39}
HACA-2/TMSA Zefirov PC (all)	D_{40}
Number of bromine atoms	D_{41}

NB, for details of the descriptors, see (http://www.codessa-pro.com/manual/manual.htm).
PPSA, partial positive surface area; TMSA, total molecular surface area; HDSA-2, area weighted surface charge of hydrogen bonding donor atoms; HACA-2, area weighted surface charge of hydrogen bonding acceptor atoms; and HDCA-1, hydrogen bonding donor ability of the molecule.

A model involving both molecular and fragment descriptors was found for Log(*Rf*:*a*), although its statistical parameters were similar to those for the model involving only CODESSA-PRO descriptors (Eq. 23, Table 1). The plot of correlation coefficients versus the number of descriptors is shown in SM2 in Fig. 1l.

4.6. Discussion of the proposed QSAR models

Overall, the correlation coefficients for the human blood:air and human and rat tissue:air partition coefficients are well correlated using the POOL 1 descriptor set. The correlation coefficients for Eqs. 1–11 range from $R^2 = 0.881$ to $R^2 = 0.983$. A combination of the POOL 1 and POOL 2 descriptors did not show any significant improvement in statistical characteristics.

A physical interpretation of the descriptors involved in the developed regression models can be based on Tomasi and Persico[121] and Karelson[122] approaches for solute–solvent interactions.

According to a commonly accepted classification of solute–solvent interactions[122,123] practically all molecular descriptors appearing in the model Eqs. 1–11, 23 belong to one of the following categories:

(i) The descriptors (D_1–D_{11}, D_{33}, D_{38}, D_{39}, and D_{41}) reflect and various aspects of the size and shape of the molecule and can be related to *cavity formation* in the condensed medium (essentially a concentrated aqueous solution–blood, tissue).

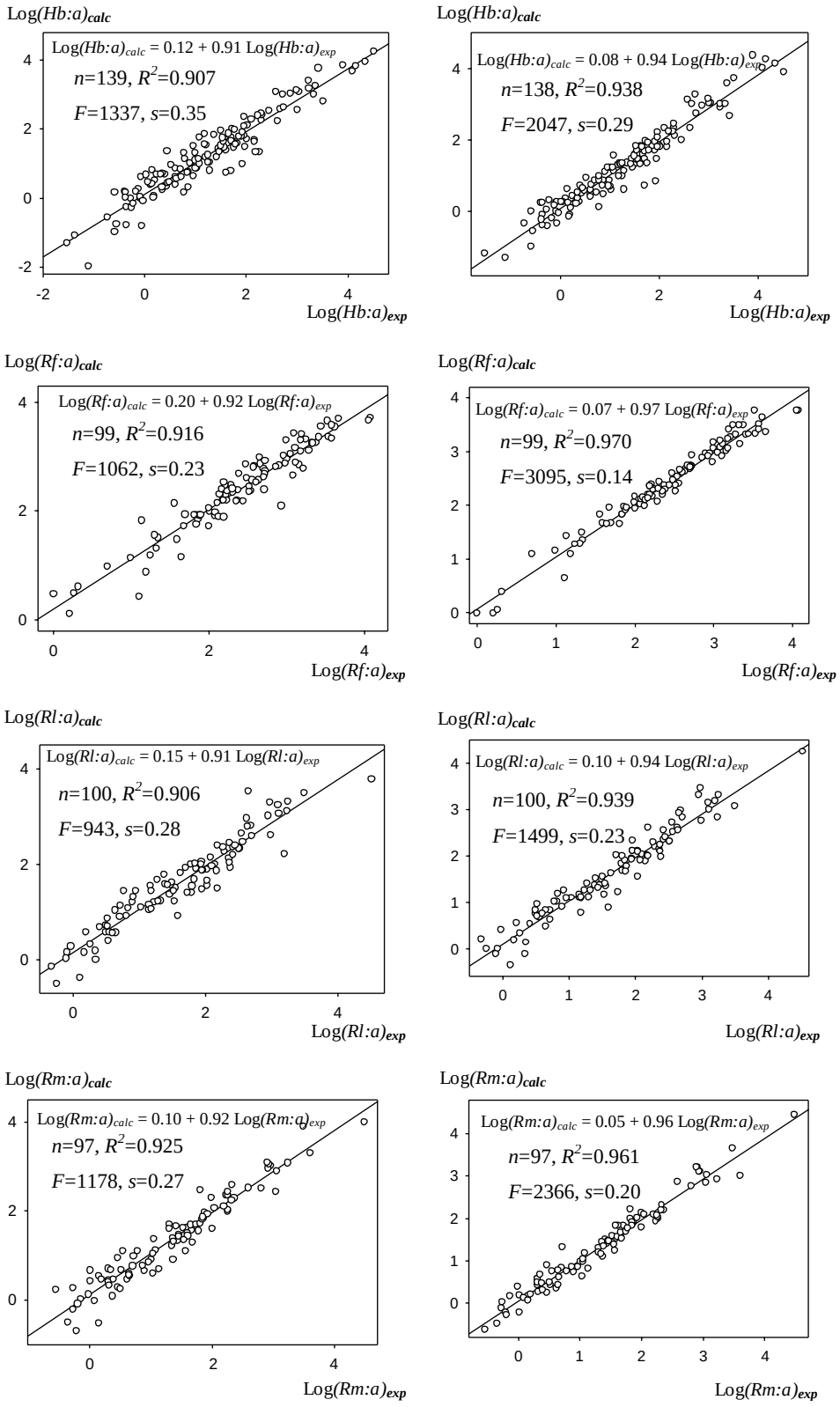
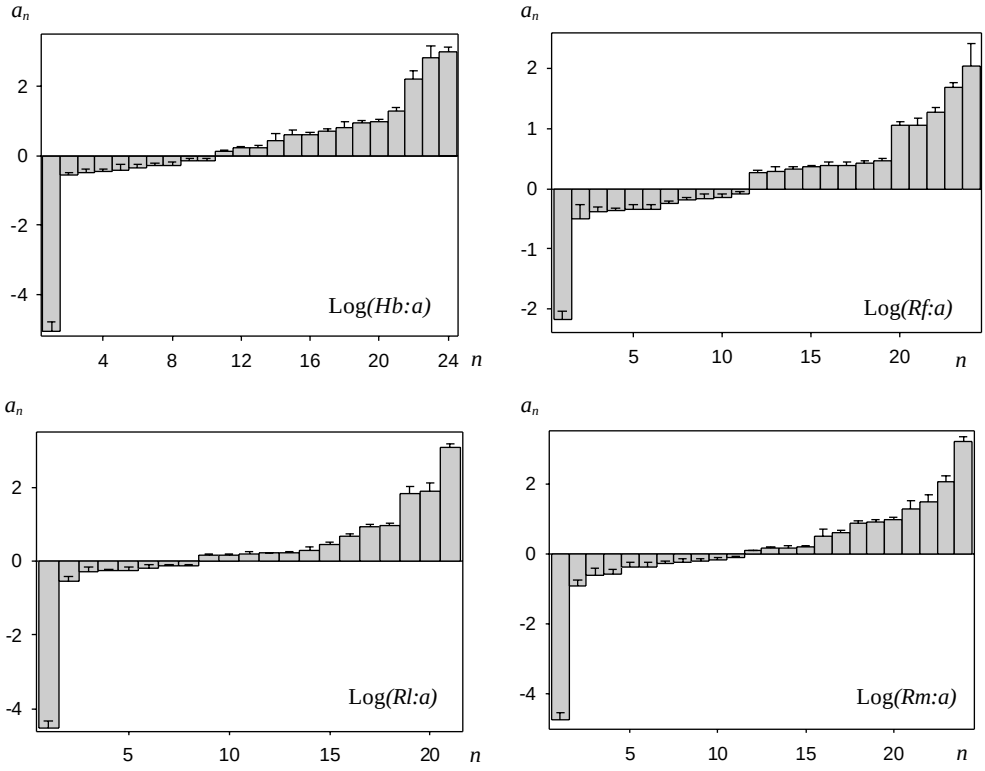


Figure 2. CODESSA-PRO (left) and ISIDA (right) modeling: plots of calculated versus observed values for four largest data sets: Log(*Hb*:*a*), Log(*Rf*:*a*), Log(*Rl*:*a*), and Log(*Rl*:*a*).

(ii) The descriptors (D_{12}–D_{15}, D_{17}–D_{19}, D_{32}, and D_{34}–D_{36}) can be related to *electrostatic interactions* between the solute and solvent molecules, and characterize the charge distribution in the solute molecules.

(iii) The descriptors (D_{25}–D_{29}) can be related to *dispersion interactions* between the solute and solvent molecules, and characterize the electron orbital distribution and energetics.



n	Log(*Hb*:*a*)	Log(*Rf*:*a*)	Log(*Rl*:*a*)	Log(*Rm*:*a*)	n	Log(*Hb*:*a*)	Log(*Rf*:*a*)	Log(*Rl*:*a*)	Log(*Rm*:*a*)
1	C—O—C	C—O—C	C—O—C	C—O—C	13	C—C	F—C—F	C—C*C	C—C=C
2	F—C—F	C—C—N	O—C—F	O—C—F	14	C=C·C	C·C	C·C·O	C—C*C
3	O—C—F	Cl—C*C	Cl—C—Br	F—C—F	15	C—F	C*C	C·O	C*C
4	Cl—C—F	Cl—C—Cl	Cl—C—Cl	Cl—C—F	16	Cl—C	a_0[a]	Cl—C	C—F
5	O—C=O	Cl—C—Br	C—C—O	C=C—F	17	Br—C	C—C=O	C—C=O	Cl—C
6	a_0[a]	C—F	Cl—C—F	O—C=O	18	C:C	C—C	Br—C	CO
7	Cl—C—Cl	C—C—O	Cl—C=C	Cl—C—Cl	19	C·O	C·O	C—N	Br—C
8	C—C—F	Cl—C—C	F—C—F	Cl—C—Br	20	C—C=O	Cl—C	C≡N	C—C=O
9	C—C·C	Br—C—C	C·C	a_0[a]	21	C·C=O	C≡N	C—O	C:C
10	C—C—C	Cl—C=C	C—C—C	C—C—O	22	C—N	Br—C		C—N
11	C·C	C—C—C	C—C=C	Cl—C=C	23	C≡N	C—O		C≡N
12	C*C	C—C=C	C*C	C—C	24	C—O	C—N		C—O

[a]The a_o term is fragment independent.

Figure 3. Contributions (a_n) of the fragments to partition coefficients Log(*Hb*:*a*), Log(*Rf*:*a*), Log(*Rl*:*a*), and Log(*Rm*:*a*) for the ISIDA models obtained for the full data sets. The standard deviations are given as error bars. C—C, C·C, C=O, C:C, C≡N, and C*C are molecular fragments with single (in chain and in cycle) double (in chain and in cycle), triple, and aromatic bonds, respectively. The fragment numbers (*n*) correspond to the following fragments:

(iv) The descriptors (D_{16}, D_{17}, D_{20}–D_{23}, D_{33}, D_{37}, and D_{40}) can be related to *specific* (*hydrogen bonding*) *interactions* between the solute and solvent molecules, and characterize the charge distribution on the potentially hydrogen bonded atoms. In principle, the various counts of heteroatoms and the bonds with heteroatoms (D_2–D_5, D_{38}) also belong to this class of descriptors, as the respective atoms may also potentially participate in specific intermolecular interactions.

Significantly, in each of the eleven QSAR models for the partition coefficients (Eqs. 1–11 and 23) there are descriptors from at least three of the above-listed four classes. Consequently, these 4- or 5-parameter equations represent close to the minimum possible information for a complete description of the solvation interactions determining distributions between biological solutions and the gas phase.

Plots of the predicted versus observed values for the regression models (Eqs. 1–11 and 23) are given in Figure 2 or in the supplementary material (see SM7 and SM8).

The fragment contributions (the POOL 2 descriptor set) bring useful information about the influence of particular functional groups on the biosolubility of chemicals, as illustrated by large contributions to Log(*Hb*:*a*), Log(*Rl*:*a*), and Log(*Rm*:*a*) of some groups containing two or three atoms. Thus, C—O of hydroxyl group of alcohols, C·C=O or C—C=O of cyclic or acyclic ketones, C≡N of nitriles, and C·O of cyclic ethers

Table 3. Internal validation of the QSAR models obtained with CODESSA-PRO and ISIDA approaches.[a]

Training set	N	k[b]	R^2	s^2	Test set	n_{test}[c]	R^2(pred)	s^2(pred)
CODESSA-PRO models								
Log(*Hb*:*a*)/Eq. 1								
a + b	93		0.908	0.145	C	46	0.905	0.123
a + c	93		0.920	0.148	B	46	0.879	0.183
b + c	92		0.897	0.128	A	47	0.930	0.090
Average			0.908	0.140			0.905	0.132
Log(*Hf*:*a*)/Eq. 2								
a + b	28		0.917	0.030	C	14	0.917	0.038
a + c	28		0.944	0.038	B	14	0.861	0.066
b + c	28		0.916	0.053	A	14	0.917	0.031
Average			0.926	0.040			0.898	0.045
Log(*Hbr*:*a*)/Eq. 3								
a + b	24		0.968	0.043	C	12	0.918	0.072
a + c	24		0.960	0.049	B	12	0.963	0.030
b + c	24		0.960	0.035	A	12	0.960	0.056
Average			0.963	0.042			0.947	0.053
Log(*Hl*:*a*)/Eq. 4								
a + b	20		0.902	0.022	C	10	0.829	0.082
a + c	20		0.897	0.047	B	10	0.760	0.046
b + c	20		0.886	0.059	A	10	0.905	0.024
Average			0.895	0.043			0.831	0.051
Log(*Hm*:*a*)/Eq. 5								
a + b	26		0.947	0.066	C	13	0.857	0.116
a + c	26		0.926	0.102	B	13	0.872	0.091
b + c	26		0.872	0.113	A	13	0.957	0.058
Average			0.915	0.094			0.895	0.088
Log(*Hk*:*a*)/Eq. 6								
a + b	23		0.951	0.050	C	11	0.963	0.053
a + c	23		0.951	0.057	B	11	0.955	0.041
b + c	22		0.962	0.049	A	12	0.936	0.054
Average			0.955	0.052			0.951	0.049
Log(*Rf*:*a*)/Eq. 7								
a + b	66		0.940	0.051	C	33	0.854	0.089
a + c	66		0.888	0.068	B	33	0.952	0.039
b + c	66		0.917	0.069	A	33	0.916	0.050
Average			0.915	0.063			0.907	0.059
Log(*Rbr*:*a*)/Eq. 8								
a + b	40		0.917	0.059	C	19	0.875	0.069
a + c	39		0.906	0.072	B	20	0.908	0.031
b + c	39		0.894	0.058	A	20	0.918	0.068
Average			0.906	0.063			0.900	0.056
Log(*Rl*:*a*)/Eq. 9								
a + b	67		0.908	0.092	C	33	0.901	0.099
a + c	67		0.894	0.103	B	33	0.935	0.069
b + c	66		0.922	0.084	A	34	0.871	0.115
Average			0.908	0.093			0.902	0.094
Log(*Rm*:*a*)/Eq. 10								
a + b	65		0.917	0.080	C	32	0.935	0.087
a + c	65		0.923	0.093	B	32	0.927	0.063
b + c	64		0.937	0.078	A	33	0.892	0.087
Average			0.926	0.084			0.918	0.079
Log(*Rk*:*a*)/Eq. 11								
a + b	18		0.991	0.008	C	9	0.959	0.022
a + c	18		0.986	0.013	B	9	0.975	0.011
b + c	18		0.987	0.007	A	9	0.933	0.116
Average			0.988	0.009			0.956	0.050
Log(*Rf*:*a*)/Eq. 23								
a + b	66		0.925	0.064	C	33	0.906	0.058
a + c	66		0.891	0.067	B	33	0.9536	0.033
b + c	66		0.938	0.051	A	33	0.8754	0.083
Average			0.918	0.061			0.912	0.058

(*continued on next page*)

Table 3 (*continued*)

Training set	N	k[b]	R^2	s^2	Test set	n_{test}[c]	R^2(pred)	s^2(pred)
ISIDA models								
Log(*Hb*:*a*)/Eq. FE1								
a + b	92	21	0.945	0.10	C	42	0.869	0.19
a + c	92	22	0.937	0.12	B	44	0.908	0.15
b + c	92	19	0.938	0.10	A	46	0.880	0.16
Average			0.940	0.11			0.886	0.17
Log(*Rf*:*a*)/Eq. FE1								
a + b	66	23	0.983	0.020	C	31	0.894	0.068
a + c	66	22	0.965	0.029	B	32	0.948	0.048
b + c	66	23	0.972	0.032	A	31	0.938	0.044
Average			0.973	0.026			0.927	0.053
Log(*Rl*:*a*)/Eq. FE1								
a + b	67	17	0.936	0.078	C	30	0.843	0.15
a + c	67	17	0.932	0.078	B	32	0.906	0.090
b + c	66	17	0.946	0.068	A	33	0.843	0.15
Average			0.938	0.073			0.864	0.13
Log(*Rm*:*a*)/Eq. FE1								
a + b	65	21	0.959	0.053	C	32	0.922	0.11
a + c	65	18	0.955	0.068	B	32	0.859	0.11
b + c	64	17	0.966	0.053	A	29	0.882	0.12
Average			0.960	0.058			0.888	0.12

[a] See text for the preparation of training and test sets. Statistical parameters for the sets: the number of compounds in the training set (n), correlation coefficient (R), and standard deviation (s) for the model; the number of compounds in the test set (n_{test}), correlation coefficient R(pred), and standard deviation s(pred) for the linear correlation $\mathrm{Log}(X\mathrm{pred}) = a\,\mathrm{Log}(X\mathrm{exp}) + b\,(X = H, R)$ between experimental and predicted partition coefficients.

[b] k is the number of fitted coefficients (fragment contributions).

[c] Some compounds (from 0 to 4) can be excluded from the test set if they contain unique molecular fragment(s) which were not presents in the corresponding training set.

(Fig. 3). The C—Br, C—Cl of polybromo- and polychloroalkanes, C—C of alkanes, and C*C of aromatic compounds contribute significantly to Log(*Rf*:*a*) (Fig. 3). The fragment contributions can be used for the preparation of focused virtual combinatorial libraries of chemicals and potential drugs.[66,123]

4.7. Internal validation of QSAR models

An important aspect of any QSAR study is the validation of the model. For internal validation, the parent data set was divided into three subsets: the first, fourth, seventh, etc., entries form the first subset (#1), the second, fifth, eighth, etc., entries form the second subset (#2), and the third, sixth, ninth, etc., form the third subset (#3). Three training sets were prepared as a combination of two subsets, Set I, (#1 and #2). Set II (#1 and #3), and Set III (#2 and #3). For each training set, the remaining subsets (#3, #2, and #1, respectively) constituted the test set.

For each training set, the correlation equation was developed with the same set of descriptors obtained for the whole data set. The obtained equation was then used to predict the property values for the compounds from the corresponding test set. The efficiency of the QSAR models was estimated using the above internal validation methodology, to determine the correlation both for the full set and for each training set. Correlation coefficients (R(pred)) and standard deviations (s(pred)) of linear correlations between experimental values of blood:air and tissue:air partition coefficients and those predicted for the test sets were also calculated.

The validity of the proposed 4 and 5 descriptor regression models Eqs. 1–11 and 23 was also tested using the internal validation methodology[48]. The results show that the predicted correlation coefficient values are in good agreement with our original regression models (see Table 3), where n is the number of data points, R^2 is the squared correlation coefficient and s is the standard deviation. The statistical characteristics R and s are given for both the fitted equation and for the predicted data (pred).

'Prediction' calculations with ISIDA have been done for the four largest sets Log(*Hb*:*a*), Log(*Rf*:*a*), Log(*Rl*:*a*), and Log(*Rm*:*a*) involving, respectively, 138, 99, 100, and 97 data points. As seen from Table 3, models selected at the training stage predict well the properties of the test set compounds: the mean (over three test sets) statistical criteria for the linear correlation calculated versus experimental values are $R^2 = 0.864$–0.927; $s^2 = 0.053$–0.17; which is slightly better than those obtained with CODESSA-PRO (Table 3).

5. Conclusions

The prediction of partition coefficient is an important indicator in the determination of blood and

tissue solubilities in humans and rats. We have developed QSAR models for the prediction of human blood:air partition coefficient, as well as human and rat tissue:air partition coefficients, using the descriptors solely calculated from structures. As the data set consists of diverse compounds, the models can be used reliably to predict biological partition coefficients for a wide range of chemical structures, including those not yet synthesized. Based on the above observations, we expect that both approaches will also provide satisfactory correlation equations for predicting tissue:blood partition coefficients (such as brain:blood partition coefficients) for both human and rat.

Acknowledgments

V.S. thanks the French Embassy in Russia for supporting his stage in the Laboratory of Chemoinformatics at the Louis Pasteur University of Strasbourg in 2004. The Estonian Science Foundation (Grant No. 4548) is gratefully acknowledged for the partial support of this work.

Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.bmc.2005.06.066.

References and notes

1. Weathersby, P. K.; Homer, L. D. *Undearsea Biomed. Res.* **1980**, *7*, 277.
2. Fiserova-Bergerova, V. *Model Inhalation Exposure Vap.* **1983**, *1*, 3.
3. Sato, A.; Nakajima, T. *Br. J. Ind. Med.* **1979**, *36*, 231.
4. Gargas, M. L.; Burgess, R. J.; Voisard, D. E.; Cason, G. H.; Andersen, M. E. *Toxicol. Appl. Pharmacol.* **1989**, *98*, 87.
5. Fiserova Bergerova, V.; Tichy, M.; Di Carlo, F. J. *Drug Metab. Rev.* **1984**, *15*, 1033.
6. Fiserova-Bergerova, V.; Diaz, M. L. *Int. Arch. Occup. Environ. Health* **1986**, *58*, 75.
7. Andersen, M. E. *Drug Metab. Rev.* **1982**, *13*, 799.
8. Andersen, M. E. *Model Inhalation Exposure Vap.* **1983**, *2*, 67.
9. Ramsey, J. C.; Andersen, M. E. *Toxicol. Appl. Pharmacol.* **1984**, *73*, 159.
10. Gargas, M. L.; Andersen, M. E.; Clewell, H. J., III *Toxicol. Appl. Pharmacol.* **1986**, *86*, 341.
11. Gargas, M. L.; Clewell, H. J., III; Andersen, M. E. *Toxicol. Appl. Pharmacol.* **1986**, *82*, 211.
12. D'Souza, R. W.; Francis, W. R.; Andersen, M. E. *J. Pharmacol. Exp. Ther.* **1988**, *245*, 563.
13. Hetrick, D. M.; Jarabek, A. M.; Travis, C. C. *J. Pharmacol. Biopharm.* **1991**, *19*, 1.
14. Loizou, G. D.; Anders, M. W. *Drug Metab. Dispos.* **1993**, *21*, 634.
15. Loizou, G. D.; Urban, G.; Dekant, W.; Anders, M. W. *Drug Metab. Dispos.* **1994**, *22*, 511.
16. Gearhart, J. M.; Seckel, C.; Vinegar, A. *Toxicol. Appl. Pharmacol.* **1993**, *119*, 258.
17. Loizou, G. D.; Anders, M. W. *Drug Metabol. Dispos.* **1995**, *23*, 875.
18. Vinegar, A.; Williams, R. J.; Fisher, J. W.; McDougal, J. N. *Toxicol. Appl. Pharmacol.* **1994**, *129*, 103.
19. Evans, M. V.; Crank, W. D.; Yang, H. M.; Simmons, J. E. *Toxicol. Appl. Pharmacol.* **1994**, *128*, 36.
20. Ellis, M. K.; Trebilcock, R.; Naylor, J. L.; Tseung, K.; Collins, M. A.; Hext, P. M.; Green, T. *Fundam. Appl. Toxicol.* **1996**, *31*, 243.
21. Kedderis, G. L.; Carfagna, M. A.; Held, S. D.; Batra, R.; Murphy, J. E.; Gargas, M. L. *Toxicol. Appl. Pharmacol.* **1993**, *123*, 274.
22. Csanady, G. A.; Kessler, W.; Hoffmann, H. D.; Filser, J. G. *Toxicol. Lett.* **2003**, *138*, 75.
23. Nihlen, A.; Johanson, G. *Toxicol. Sci.* **1999**, *51*, 184.
24. Tardif, R.; Charest-Tardif, G.; Brodeur, J.; Krishnan, K. *Toxicol. Appl. Pharmacol.* **1997**, *144*, 120.
25. Plowchalk, D. R.; Andersen, M. E.; Bogdanffy, M. S. *Toxicol. Appl. Pharmacol.* **1997**, *142*, 386.
26. Cantoreggi, S.; Keller, D. A. *Toxicol. Appl. Pharm.* **1997**, *143*, 130.
27. Kirman, C.; Gargas, M. L.; Deskin, R.; Tonner-Navarro, L.; Andersen, M. *J. Toxicol. Environ. Health A* **2003**, *66*, 253.
28. Sweeney, L. M.; Gargas, M. L.; Strother, D. E.; Kedderis, G. L. *Toxicol. Sci.* **2003**, *71*, 27.
29. Hayes, S. M.; Elswick, B. A.; Blunenthal, G. M.; Welsch, F.; Conolly, R. B.; Gargas, M. L. *Toxicol. Appl. Pharmacol.* **2000**, *163*, 67.
30. Gargas, M. L.; Tyler, T. R.; Sweeney, L. M.; Corley, R. A.; Weitz, K. K.; Mast, T. J.; Paustenbach, D. J.; Hays, S. M. *Toxicol. Appl. Pharmacol.* **2000**, *165*, 53.
31. Csanady, G. A.; Denk, B.; Krezeur, P. E.; Kessler, W.; Baur, C.; Gargas, M. L.; Fisler, J. G. *Toxicol. Appl. Pharmacol.* **2000**, *165*, 1.
32. Barton, H. A.; Creech, J. R.; Godin, C. S.; Randall, G. M.; Seckel, C. S. *Toxicol. Appl. Pharmacol.* **1995**, *130*, 237.
33. Reitz, R. H.; McCroskey, P. S.; Park, C. N.; Andersen, M. E.; Gargas, M. L. *Toxicol. Appl. Pharmacol.* **1990**, *105*, 37.
34. Kaneko, T.; Horiuchi, J.; Sato, A. *Pharmacol. Res.* **2000**, *42*, 465.
35. Andersen, M. E.; Clewell, H. J.; Mahle, D. A.; Gearhart, J. M. *Toxicol. Appl. Pharmacol.* **1994**, *128*, 158.
36. Payne, M. P.; Kenny, L. C. *J. Toxicol. Environ. Health A* **2002**, *65*, 897.
37. Meulenberg, C. J. W.; Vijverberg, H. P. M. *Toxicol. Appl. Pharmacol.* **2000**, *165*, 206.
38. Abraham, M. H.; Weathersby, P. K. *J. Pharm. Sci.* **1994**, *83*, 1450.
39. Fouchecourt, M. O.; Beliveau, M.; Krishnan, K. *Sci. Total Environ.* **2001**, *274*, 125.
40. Poulin, P.; Krishnan, K. *Toxicol. Methods* **1996**, *6*, 117.
41. Pelekis, M.; Krewski, D.; Krishnan, K. *Toxicol. Methods* **1997**, *7*, 205.
42. Poulin, P.; Krishnan, K. *Int. J. Toxicol.* **1999**, *18*, 7.
43. Basak, S. C.; Mills, D.; Hawkins, D. M.; El-Masri, H. A. *SAR QSAR Environ. Res.* **2002**, *13*, 649.
44. Basak, S. C.; Mills, D.; Hawkins, D. M.; El-Masri, H. A. *Risk Anal.* **2003**, *23*, 1173.
45. Basak, S. C.; Mills, D.; El-Masri, H. A.; Mumtaz, M. M.; Hawkins, D. M. *Environ. Toxicol. Pharmacol.* **2004**, *16*, 45.
46. Béliveau, M.; Lipscomb, J.; Tardif, R.; Krishnan, K. *Chem. Res. Toxicol.* **2005**, *18*, 475.
47. Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162.
48. Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720.

49. Katritzky, A. R.; Kuanar, M.; Fara, D. C.; Karelson, M., Jr.; Acree, W. E., Jr. *Bioorg. Med. Chem.* **2004**, *12*, 4735.
50. CODESSA-PRO *User's Manual*, http://www.codessa-pro.com/manual/manual.htm.;
51. Katritzky, A. R.; Fara, D. C.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. *Curr. Top. Med. Chem.* **2002**, *12*, 1333.
52. Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 529.
53. Katritzky, A. R.; Fara, D. C.; Karelson, M. *Bioorg. Med. Chem.* **2004**, *12*, 3027.
54. Trepalin, S. V.; Gerasimenko, V. A.; Kozyukov, A. V.; Savchuk, N.; Ivaschenko, A. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 249.
55. Klopman, G.; Tu, M. *J. Med. Chem.* **1999**, *42*, 992.
56. Hansch, C.; Hoekman, D.; Leo, A.; Zhang, L.; Li, P. *Toxicol. Lett.* **1995**, *79*, 45.
57. Zefirov, N. S.; Palyulin, V. A. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1112.
58. Anzali, S.; Barnickel, G.; Cezanne, B.; Krug, M.; Filimonov, D.; Poroikov, V. *J. Med. Chem.* **2001**, *44*, 2432.
59. Klopman, G.; Zhu, H. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 439.
60. Avidon, V. V. *Khimiko-Farmatsevticheskii Zhurnal* **1974**, *8*, 22.
61. Bawden, D. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 14.
62. Poroikov, V. V.; Filimonov, D. A.; Borodina Yu, V.; Lagunin, A. A.; Kos, A. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1349.
63. Solov'ev, V. P.; Varnek, A.; Wipff, G. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 847.
64. Varnek, A.; Wipff, G.; Solov'ev, V. P. *Solvent Extraction and Ion Exchange* **2001**, *19*, 791.
65. Varnek, A.; Wipff, G.; Solov'ev, V. P.; Solotnov, A. F. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 812.
66. Solov'ev, V. P.; Varnek, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1703.
67. ISIDA (In Silico Design and Data Analysis) informational system http://infochim.u-strasbg.fr/recherche/isida/index.php.;
68. Tichy, M.; Fiserova-Bergerova, V.; Di Carlo, F. J. *Pharmochem. Libr.* **1985**, *8*, 225.
69. Imbriani, M.; Ghittori, S.; Pezzagno, G.; Capodaglio, E. G. *Ital. Med. Lav.* **1985**, *7*, 133.
70. Perbellini, L.; Brugnone, F.; Caretta, D.; Maranelli, G. *Br. J. Ind. Med.* **1985**, *42*, 162.
71. Fisher, J.; Mahle, D.; Bankston, L.; Greene, R.; Gearhart, J. *Am. Ind. Hyg. Assoc. J.* **1997**, *58*, 425.
72. Cowles, A. L.; Borgstedt, H. H.; Gillies, A. J. *Anesthesiology* **1971**, *35*, 203.
73. Steward, A.; Allott, P. R.; Cowles, A. L.; Mapleson, W. W. *Br. J. Anaesthesia* **1973**, *45*, 282.
74. Droz, P. O.; Berode, M.; Jang, J. Y. *Am. Ind. Hyg. Assoc. J.* **1999**, *60*, 243.
75. Filser, J. G.; Schmidbauer, R.; Rampf, F.; Baur, C. M.; Putz, C.; Csanady, G. A. *Toxicol. Appl. Pharm.* **2000**, *169*, 40.
76. Lin, Y. S.; Smith, T. J.; Kelsey, K. T.; Wypij, D. *Environ. Health Perspect.* **2001**, *109*, 921.
77. Falk, A.; Gullstrand, E.; Loef, A.; Wigaeus-Hjelm, E. *Br. J. Ind. Med.* **1990**, *47*, 62.
78. Jaernberg, J.; Johanson, G. *Toxicol. Ind. Health* **1995**, *11*, 81.
79. Sato, A.; Nakajima, T. *Arch. Environ. Health* **1979**, *34*, 69.
80. Batterman, S.; Zhang, L.; Wang, S.; Franzblau, A. *Sci. Total Environ.* **2002**, *284*, 237.
81. Gatley, S. J.; Hichwa, R. D.; Shaughnessy, W. J.; Nickles, R. J. *Int. J. Appl. Radiat. Isotop.* **1981**, *32*, 211.
82. Eger, E. I., II; Liu, J.; Koblin, D. D.; Laster, M. J.; Taheri, S.; Halsey, M. J.; Ionescu, P.; Chortkoff, B. S.; Hudlicky, T. *Anesth. Analg.* **1994**, *79*, 245.
83. Franks, P. J.; Hooper, R. H.; Jones, P. R. M. *Br. J. Anaesth.* **1989**, *62*, 425.
84. Varene, N.; Choukroun, M. L.; Marthan, R.; Varene, P. *J. Appl. Physiol.* **1989**, *66*, 2468.
85. Niazi, S.; Chiou, W. L. *J. Pharm. Sci.* **1975**, *64*, 1538.
86. Meyer, M.; Scheid, P. *J. Appl. Physiol.* **1980**, *48*, 1035.
87. Eger, R. R.; Eger, E. I., II *Anesth. Analg.* **1985**, *64*, 640.
88. Kaneko, T.; Wang, P. Y.; Sato, A. *Occup. Environ. Med.* **1994**, *51*, 68.
89. Nihlen, A.; Loef, A.; Johanson, G. *J. Expos. Anal. Environ. Epidem.* **1995**, *5*, 573.
90. Csanady, G. A.; Kreuzer, P. E.; Baur, C.; Filser, J. G. *Toxicology* **1996**, *113*, 300.
91. Csanady, G. A.; Filser, J. G. *Arch. Toxicol.* **2001**, *74*, 663.
92. Johanson, G.; Dynesius, B. *Br. J. Ind. Med.* **1988**, *45*, 561.
93. Eger, E. I., II *Anesth. Analg.* **1987**, *66*, 971.
94. Yasuda, N.; Trag, A. G.; Eger, E. I., II *Anesth. Analg.* **1989**, *69*, 370.
95. Loizou, G. D.; Eldirdiri, N. I.; King, L. J. *Inhal. Toxicol.* **1996**, *8*, 1.
96. http://www.epa.gov/appt/chemtest/vch.pdf; viewed December 12, 2003.
97. Thrall, K. D.; Gies, R. A.; Muniz, J.; Woodstock, A. D.; Higgins, G. *J. Toxicol. Environ. Health A* **2002**, *65*, 2075.
98. Meulenberg, C. J. W.; Wijnker, A. G.; Vjverberg, H. P. M. *J. Toxicol. Environ. Health A* **2003**, *66*, 1985.
99. Kumarathasan, P.; Otson, R.; Chu, I. *Chemosphere* **1998**, *37*, 159.
100. Simmons, J. E.; Boyes, W. K.; Bushnell, P. J.; Raymer, J. H.; Limsakun, T.; McDonald, A.; Sey, Y. M.; Evans, M. V. *Toxicol. Sci.* **2002**, *69*, 3.
101. Kaneko, T.; Kim, H. Y.; Wang, P. Y.; Sato, A. *J. Occup. Health* **1997**, *39*, 341.
102. Medinsky, M. A.; Bechtold, W. E.; Birnbaum, L. S.; Chico, D. M.; Gerlach, R. F.; Henderson, R. F. *Fundam. Appl. Toxicol.* **1988**, *11*, 250.
103. Williams, R. J.; Venegar, A.; McDougal, J. M.; Jarabek, A. M.; Fisher, J. W. *Fundam. Appl. Toxicol.* **1996**, *30*, 55.
104. Knaak, J. B.; Smith, L. W.; Fitzpatrick, R. D.; Olson, J. R.; Newton, P. E. *Inhal. Toxicol.* **1998**, *10*, 65.
105. Lilly, P. D.; Andersen, M. E.; Ross, T. M.; Pegram, R. A. *Toxicol. Appl. Pharm.* **1998**, *150*, 205.
106. Lilly, P. D.; Andersen, M. E.; Ross, T. M.; Pegram, R. A. *Toxicology* **1997**, *124*, 141.
107. Fassoulaki, A.; Eger, E. I., II *Br. J. Anaesth.* **1986**, *58*, 327.
108. Loizou, G. D.; Tran, C. L.; Anders, M. W. *Xenobiotica* **1997**, *27*, 87.
109. Kaneko, T.; Wang, P. Y.; Sato, A. *J. Occup. Health* **2000**, *42*, 86.
110. Borghoff, S. J.; Murphy, J. E.; Medinsky, M. A. *Fundam. Appl. Toxicol.* **1996**, *30*, 264.
111. Teo, S. K. O.; Kedderis, G. L.; Gargas, M. L. *Toxicol. Appl. Pharm.* **1994**, *128*, 92.
112. Krishnan, K.; Gargas, M. L.; Fennell, T. R.; Andersen, M. E. *Toxicol. Ind. Health* **1992**, *8*, 121.
113. Johanson, G.; Filser, J. G. *Arch. Toxicol.* **1993**, *67*, 151.
114. www.mdl.com.
115. www.hyper.com.
116. MOPAC 7.0.
117. Korn, G. A., Korn T. M. *Mathematical Handbook for Scientists and Engineers. Definitions, Theorems and For-*

mulas for Reference and Review; McGraw-Hill Book Co: New York, 1968.
118. Golub, G. H.; Reinsch, C. *Numer. Math.* **1970**, *14*, 403.
119. Kendall, M. G.; Stuart, A. The Advanced Theory of Statistics. London, Griffin, Vol. 1–3, 1966.
120. DIVA 2.1 program. Accelrys, 2002.
121. Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
122. Karelson, M. *Adv. Quantum Chem.* **1997**, *28*, 141.
123. Varnek, A.; Fourches, D.; Solov'ev, V. P.; Baulin, V. E.; Turanov, A. N.; Karandashev, V. K.; Fara, D.; Katritzky, A. R. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1365.