Calculating the Partition Coefficients of Organic Solvents in Octanol/Water and Octanol/Air

Miroslava A. Nedyalkova,*,† Sergio Madurga,‡ Marek Tobiszewski,§ and Vasil Simeonov‖

†Inorganic Chemistry Department, Faculty of Chemistry and Pharmacy, University of Sofia, Sofia 1164, Bulgaria
‡Departament de Ciència de Materials i Química Física and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Catalonia, Spain
§Department of Analytical Chemistry, Faculty of Chemistry, Gdańsk University of Technology (GUT), 80-233 Gdańsk, Poland
‖Analytical Chemistry Department, Faculty of Chemistry and Pharmacy, University of Sofia, Sofia 1164, Bulgaria

S *Supporting Information*

ABSTRACT: Partition coefficients define how a solute is distributed between two immiscible phases at equilibrium. The experimental estimation of partition coefficients in a complex system can be an expensive, difficult, and time-consuming process. Here a computational strategy to predict the distributions of a set of solutes in two relevant phase equilibria is presented. The octanol/water and octanol/air partition coefficients are predicted for a group of polar solvents using density functional theory (DFT) calculations in combination with a solvation model based on density (SMD) and are in excellent agreement with experimental data. Thus, the use of quantum-chemical calculations to predict partition coefficients from free energies should be a valuable alternative for unknown solvents. The obtained results indicate that the SMD continuum model in conjunction with any of the three DFT functionals (B3LYP, M06-2X, and M11) agrees with the observed experimental values. The highest correlation to experimental data for the octanol/water partition coefficients was reached by the M11 functional; for the octanol/air partition coefficient, the M06-2X functional yielded the best performance. To the best of our knowledge, this is the first computational approach for the prediction of octanol/air partition coefficients by DFT calculations, which has remarkable accuracy and precision.
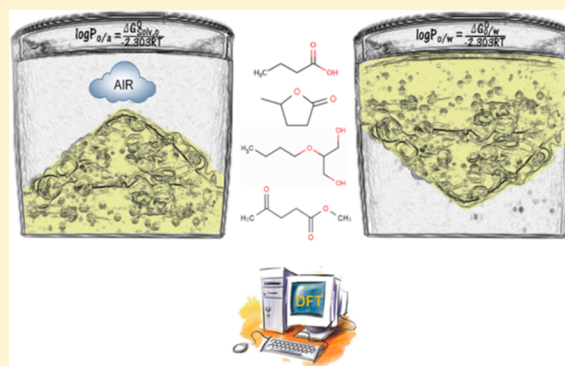
INTRODUCTION

Physical properties of molecules can be used, for example, to make predictions about the environmental fate of unknown solvents. The lack of physical property data may be resolved by the development of different computational methodologies for the prediction of the appropriate physicochemical properties. In particular, a variety of methods have been developed to predict partition coefficients from a chemical structure.[1−9] One of the most important prediction methods is based on quantitative structure−property relationships (QSPRs).[10−12] Various algorithms and online platforms based on QSPRs, such as AlogPs, ADMET predictor, and ACD/logD, have been developed. The main approach of these methods is based on finding the appropriate set of molecular descriptors that allow the precise reproduction of a given physical property using a large database of available experimental data. Accurate QSPR models are obtained when this method is applied to molecules that resemble the ones in the database used to build the model. Thus, the weakness of QSPR models is related to the prediction of properties of molecules that vary slightly from those used in the database.[13−15]

Alternatively, partition coefficients can be predicted by taking into account the fact that this property is related to the free energy difference of a solute in different solvents. In the work of Bannan et al.,[16] the computational scheme that was used consisted of molecular dynamics simulations with explicit solvent molecules to obtain transfer free energies between the solvents, which in turn were used to calculate their partition coefficients. The octanol/water and cyclohexane/water partition coefficients were obtained using the generalized AMBER force field (GAFF) and the dielectric-corrected GAFF (GAFF-DC).

Jones et al. used ab initio calculations to predict the cyclohexane/water partition coefficients for a set of 53 compounds. The free energy of transfer was calculated with several density functionals in combination with the solvation model based on density (SMD) implicit-solvent model, and a good estimation was obtained.[17] Rayne and Forest computed air−water partition coefficients for a data set of 86 large
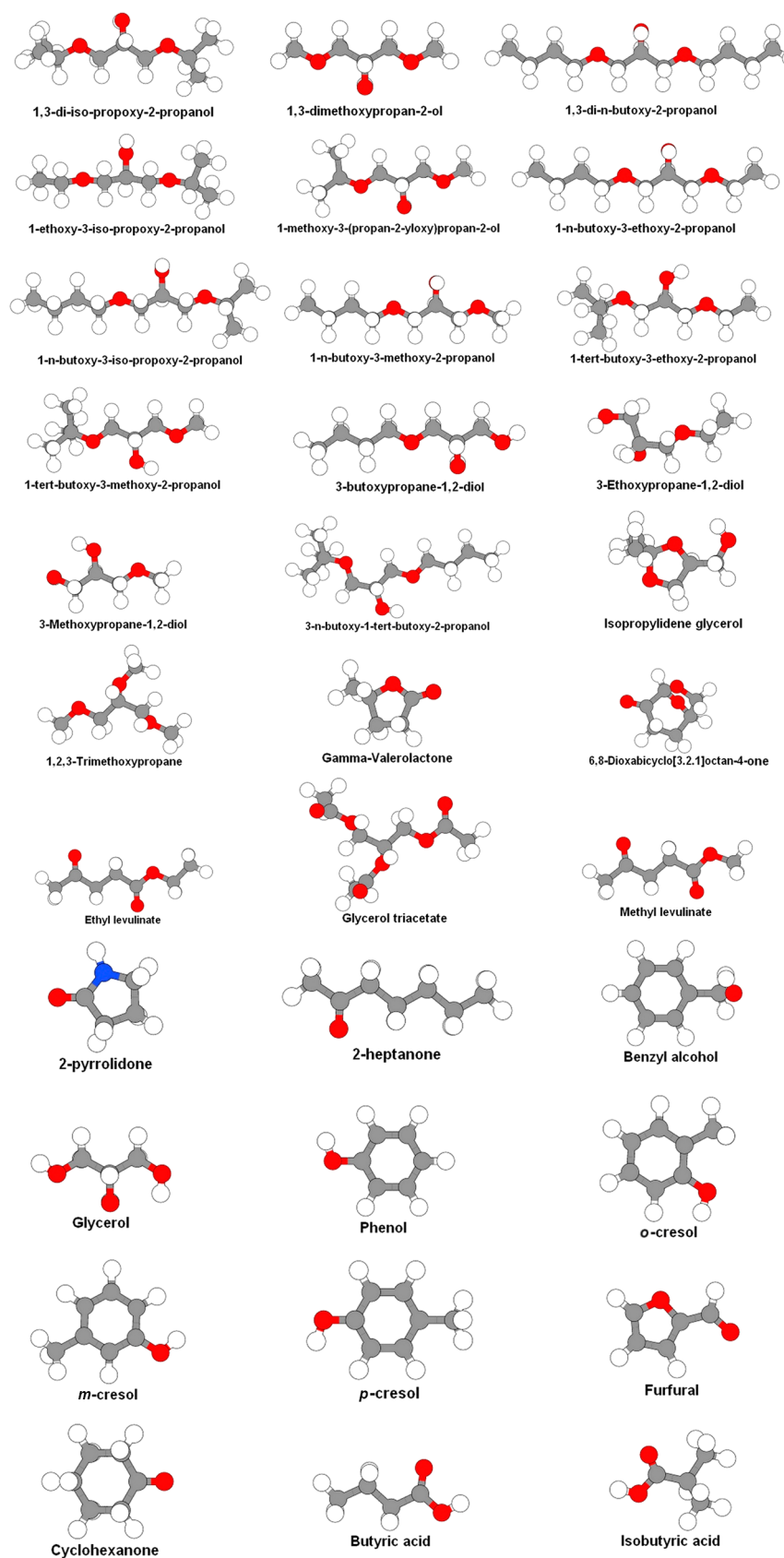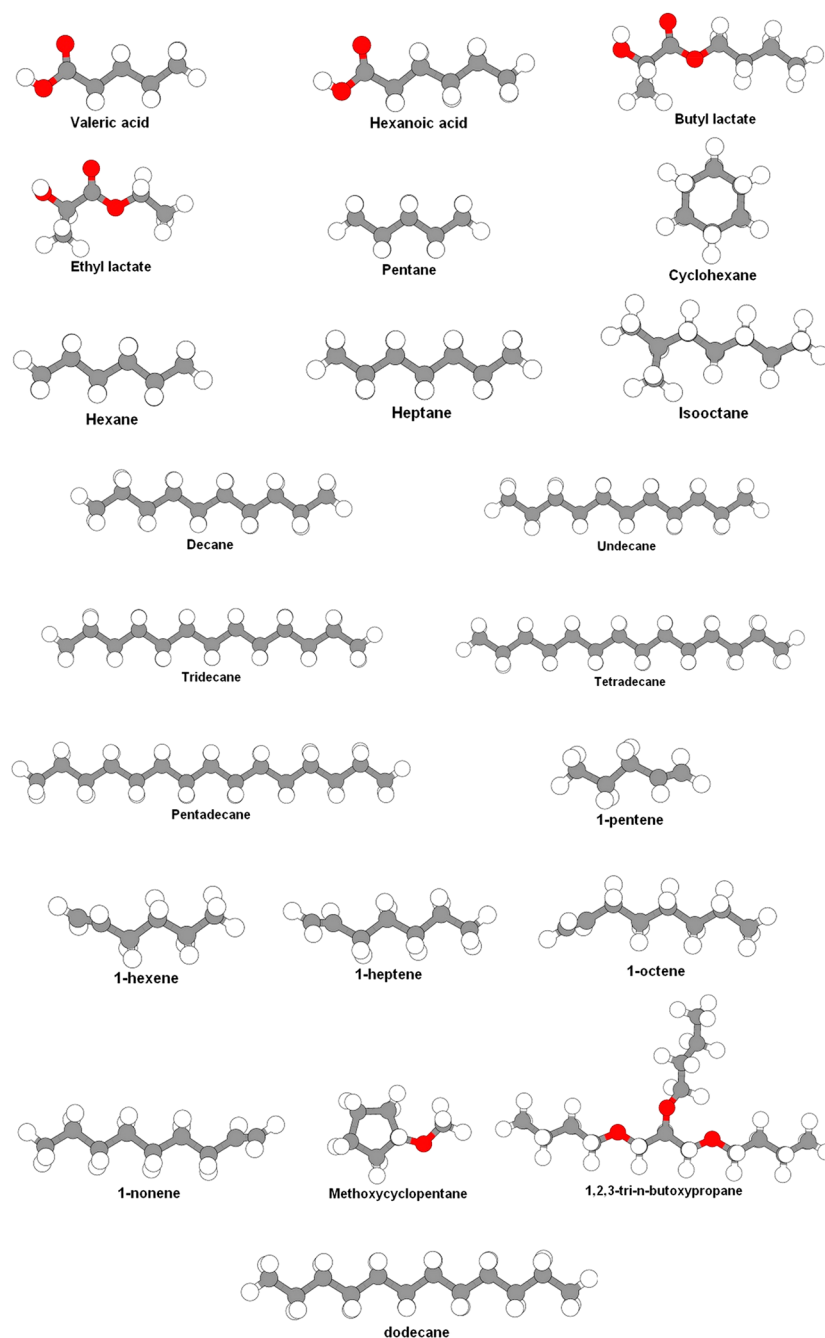
A

Figure 1. continued

Figure 1. Chemical structures of the solvents.

polycyclic aromatic hydrocarbons and their unsaturated relatives by means of high-level G4(MP2) gas- and aqueous-phase calculations with the SMD, integral equation formalism variant (IEFPCM-UFF), and conductor-like polarizable continuum model (CPCM) solvation models.[18] The results obtained using the three solvation models for a range of neutral and ionic compounds showed accurate air/water partition coefficients (K_{aw}). Better accuracy is obtained when higher levels of theory are used.

In the work of Michalík and Lukeš,[19] the octanol/water partition coefficients for 27 alkane alcohols were predicted by quantum-chemical calculations with three solvation models. The results were in rather good agreement with the corresponding experimental values. When comparing their results with those obtained using other implicit-solvent models

(IEFPCM or CPCM), the authors observed deviations from linearity. This mixed quantum (DFT)−QSPR analysis has recently been implemented successfully in the prediction of pK_a values for carboxylic acids.[20]

In the present work, the prediction of octanol/water and air/water partition coefficients for a set of 55 organic solvents was carried out by means of density functional theory (DFT) calculations. Solvation free energies were computed with various density functionals to estimate the partition coefficients. Good correlation coefficients between the calculated and experimentally measured values were obtained.

■ DATA SET

The data set consisted of 55 molecules that were selected by a previous study.[21] In that study, 150 solvents were clustered on

Table 1. Linear Regression Parameters Obtained for the Calculated LogP with Respect to the Experimental Data and Statistical Error Assessment of the Linear Regression Terms Based on Applied Computational Models

density functional	slope	intercept	R^2	R	MAD	MSE	RMSE	MAPE	standard error
M06-2X	0.9498	0.02	0.9726	0.9879	0.29	0.41	0.48	0.56	0.65
M11	0.9816	0.5398	**0.9841**	0.992	0.75	0.96	0.72	0.57	0.72
B3LYP	1.022	0.8597	0.9609	0.9802	0.86	0.96	0.9	0.58	0.86

Table 2. Linear Regression Parameters Obtained for the Calculated LogK_{oa} with Respect to the Experimental Data and Statistical Error Assessment of the Linear Regression Terms Based on Applied Computational Models

density functional	slope	intercept	R^2	R	MAD	MSE	RMSE	MAPE	standard error
M06-2X	1.0111	0.1806	**0.978**	0.9889	0.4	0.39	0.59	0.95	0.3
M11	0.9816	0.5398	0.8891	0.9428	2.18	1.45	1.2	0.9	0.6
B3LYP	0.841	0.9527	0.8618	0.9339	1.78	1.92	1.33	0.63	0.75

the basis of physicochemical properties—melting and boiling points, density, water solubility, vapor pressure, Henry's law constant, logP, logK_{oa}, and surface tension. Molecules included in the present study belong to polar or nonpolar groups. The experimental values for the octanol/water and air/water partition coefficients are presented in Table S1. The majority of these 55 molecules have been considered to be green solvents in various literature reports.[21] Some of the included molecules, such as ether−alcohols, are poorly characterized in terms of their hazards or physicochemical properties.[22]

COMPUTATIONAL METHODS

All of the calculations presented in this work were performed using the Gaussian 16 quantum chemistry package.[23] Molecular structures were generated in the more extended conformation using GaussView 5.0.[24]

The geometries of all 55 molecules were optimized using the three density functionals M06-2X,[25] M11,[26] and B3LYP[27] with the 6-311+G** basis set using the continuum solvation model based on density (SMD).[28] Hessian analysis indicated no existence of imaginary frequencies, proving that all of the optimized structures were true minima. SMD can be used as a universal solvation model because it can be applied to any charged or uncharged solute in any type of solvent.[29] The parameters required for the solvent are the dielectric constant, refractive index, bulk surface tension, and acidity and basicity parameters. This model divides the solvation free energy into two main contributions—the bulk electrostatic contribution and the cavity dispersion contribution.

To calculate the octanol/water partition coefficient, the SMD free energies obtained in the two solvents at 298.15 K were used to calculate the standard free energy associated with the transfer of the solute from the aqueous phase (w) to octanol (o):

$$\Delta G^{\circ}_{o/w} = \Delta G^{\circ}_{o} - \Delta G^{\circ}_{w} \tag{1}$$

The octanol/water partition coefficient was then calculated according to

$$\log P = -\frac{\Delta G^{\circ}_{o/w}}{2.303RT} \tag{2}$$

To calculate the octanol/air partition coefficient, the SMD solvation free energy in octanol was obtained from the free energies of the molecule in the gas phase and in octanol:

$$\Delta G^{\circ}_{solv,o} = \Delta G^{\circ}_{o} - \Delta G^{\circ}_{gas} \tag{3}$$

The octanol/air partition coefficient was then calculated according to

$$\log K_{oa} = -\frac{\Delta G^{\circ}_{solv,o}}{2.303RT} \tag{4}$$

RESULTS

The structures of the 55 solvents under study are shown in Figure 1. All of the correlation coefficients, slopes, and intercepts for all of the molecules are collected in Tables 1 and 2. The linear correlations between the experimental and calculated values for the octanol/water and octanol/air partition coefficients for the observed organic solvent data set are presented graphically in Figures 2 and 3.

Notably, a good linear correlation was obtained for the logP values with the M11 functional. As shown in Tables 1 and 2, a variety of statistical error metrics were calculated to compare the calculated logP and logK_{oa} values to the experimentally determined values: root-mean-square error (RMSE), Pearson correlation coefficient (R), mean absolute deviation (MAD), mean square error (MSE), mean absolute percentage error (MAPE), and standard error. The assumed outliers observed in the first data assessment (Figures S1−S3) were eliminated in the next level of linear regression by applying the 4σ rule[30] for the detection of outliers. Extremely large and extremely small values for both coefficients were observed in our data. The 4σ region represents 99.99% of the values for a normal distribution and 97% for symmetric unimodal distributions. In our case, the outliers were adequately identified and excluded from further statistical treatment of the data set. New plots were established and are presented in Figures 2 and 3.

These metrics allow us to make meaningful comparisons between experimental and calculated data with the SMD solvent model and the three functionals. As previously mentioned, the best results were obtained with the M11 functional, with RMSE = 0.72 and R = 0.99.

The obtained statistics show good accuracy of the computed partition coefficients with the M11 functional. In general, the signs of the calculated and experimental data were in agreement. Positive values indicate a preference for the organic phase, and negative values indicate a preference for the aqueous phase. For DFT-based SMD solvation models, this approach appears to be appropriate for obtaining good correlations with the experimental measurements in the calculation of logP. Presumably, the accuracy of the results is due to the appropriate parametrization of the SMD solvation model to yield accurate solvation free energies.
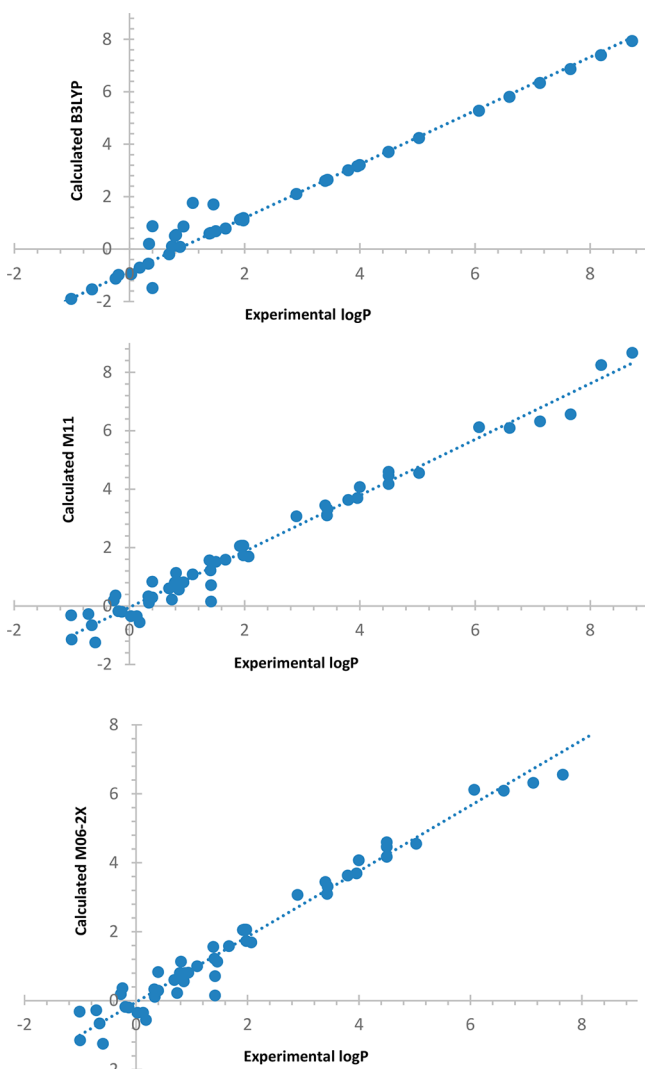
Figure 2. Comparison of the calculated and experimental logP values. The top graph shows the B3LYP model results, the middle graph shows the M11 results, and the bottom graph shows the M06-2X results.

Figure 3. Comparison of the calculated and experimental $\log K_{oa}$ values. The top graph shows the B3LYP model results, the middle graph shows the M11 results, and the bottom graph shows the M06-2X results.

The predicted values for $\log K_{oa}$ allowed us to evaluate the solvation free energies obtained with the SMD model. The various statistical error metrics were calculated and are depicted in Table 2. The best agreement between the calculated and experimental partition coefficients was obtained with the M06-2X functional, indicating a good estimation of the solvation free energy using the M06-2X functional. These findings provide useful information for understanding the partitioning, and the proposed computational scheme can be applied to other unknown solvents. For the other functionals, the obtained results also indicate that the approach is valid.

In all of the calculations, a single conformation (the more extended conformation) was used for each molecule in each phase. Although this conformation was verified to be a minimum in the gas phase and in both solvents, it is possible that multiple conformations in the gas phase and/or the solvent should be taken into account to obtain accurate descriptions.
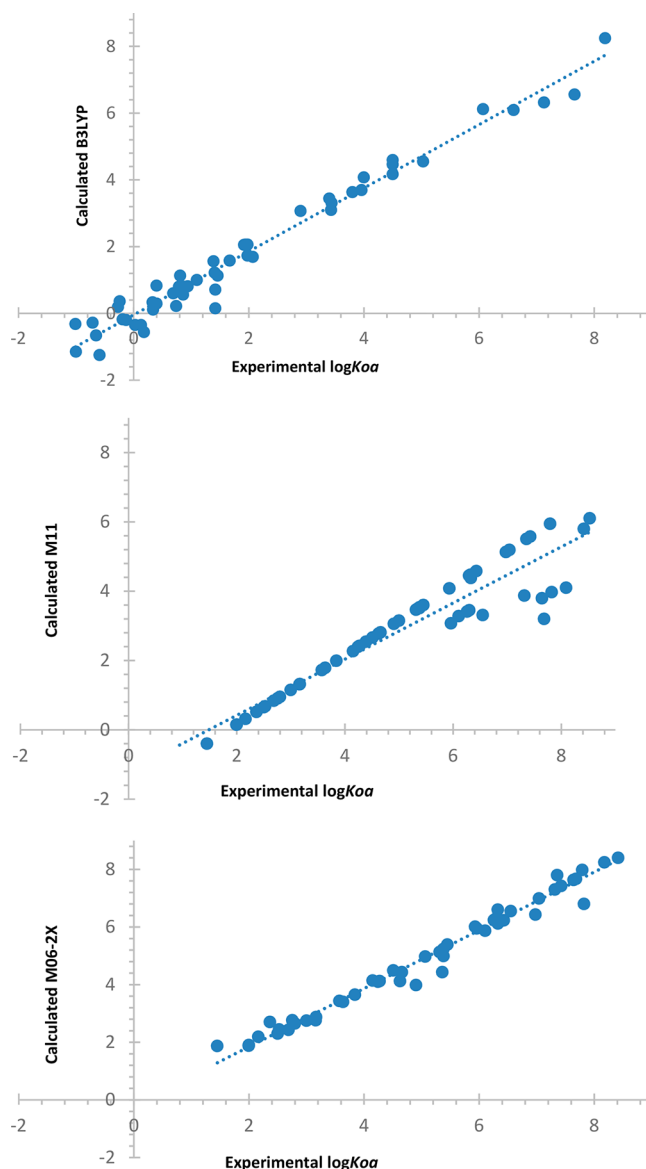
■ CONCLUSIONS

Octanol/water and octanol/air partition coefficients were predicted within the framework of DFT using three density functionals (B3LYP and two popular (but different) Minnesota meta-GGA functionals, M06-2X and M11), and their accuracies were assessed. The best correlations with the experimental data were achieved with the M11 and M06-2X functionals, respectively. The quality of the regression models was significantly improved after exclusion of points that were determined to be outliers using the 4σ rule. Thus, this computational protocol could be used as a tool for newly synthesized solvents where there is a lack of data regarding the octanol/water and octanol/air partition coefficients. The methodology should be useful for predicting missing data for envirometrics data interpretation and modeling.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00212.

Table S1 showing experimental and predicted logP and $\log K_{oa}$ values (PDF)

Formulas for evaluation metrics and Figures S1−S3 showing points deviating from linearity before application of the 4σ rule test for outliers (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: mici345@yahoo.com. Phone: 00359 885 76 33 98.

ORCID Ⓞ

Miroslava A. Nedyalkova: 0000-0003-0793-3340
Sergio Madurga: 0000-0002-8135-7057
Marek Tobiszewski: 0000-0002-9046-1649

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Kohler, M. G.; Grigoras, S.; Dunn, W. J., III. The Relationship between Chemical Structure and the Logarithm of the Partition Coefficient. *Quant. Struct.-Act. Relat.* **1988**, *7*, 150−159.

(2) Kasai, K.; Umeyama, H.; Tomonaga, A. The Study of Partition Coefficients. The Prediction of Log P Value Based on Molecular Structure. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 2701−2706.

(3) Bodor, N.; Gabanyi, Z.; Wong, C.-K. A New Method for the Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783−3786.

(4) Chou, J. T.; Jurs, P. C. Computer-Assisted Computation of Partition Coefficients from Molecular Structure Using Fragment Constants. *J. Chem. Inf. Model.* **1979**, *19*, 172−178.

(5) Sasaki, Y.; Kubodera, H.; Matuszaki, T.; Umeyama, H. Prediction of Octanol/Water Partition Coefficients Using Parameters Derived from Molecular Structures. *J. Pharmacobio-Dyn.* **1991**, *14*, 207−214.

(6) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer Automated Log P Calculation Based on an Extended Group Contribution Approach. *J. Chem. Inf. Model.* **1994**, *34*, 752−781.

(7) Meylan, W. M.; Howard, P. H. Atom/Fragment Contribution Method for Estimating Octanol−Water Partition Coefficients. *J. Pharm. Sci.* **1995**, *84*, 83−92.

(8) Ghose, A.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure−Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565−577.

(9) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley Interscience: New York, 1979; p 19.

(10) Beck, B.; Breindl, A.; Clark, T. QM/NN QSPR models with error estimation: vapor pressure and logP. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1046−1051.

(11) Šoškić, M.; Plavšić, D. Modeling the Octanol−Water Partition Coefficients by an Optimized Molecular Connectivity Index. *J. Chem. Inf. Model.* **2005**, *45*, 930−938.

(12) Duchowicz, P.; Castro, E.; Toropov, A.; Nesterov, I.; Nabiev, O. M. QSPR modeling the aqueous solubility of alcohols by optimization of correlation weights of local graph invariants. *Mol. Diversity* **2004**, *8*, 325−330.

(13) Dyekjaer, J.; Rasmussen, K.; Jonsdottir, S. QSPR models based on molecular mechanics and quantum chemical calculations. 1. Construction of Boltzmann-averaged descriptors for alkanes, alcohols, diols, ethers and cyclic compounds. *J. Mol. Model.* **2002**, *8*, 277−289.

(14) Doweyko, A. 3D-QSAR illusions. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 587−596.

(15) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. *J. Chem. Inf. Model.* **2008**, *48*, 220−232.

(16) Bannan, C.; Calabró, G.; Kyu, D.; Mobley, D. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. *J. Chem. Theory Comput.* **2016**, *12*, 4015−4024.

(17) Jones, M. R.; Brooks, B. R.; Wilson, A. K. Partition Coefficients for the SAMPL5 Challenge Using Transfer Free Energies. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1129−1138.

(18) Rayne, S.; Forest, K. Air-water partition coefficients for a suite of polycyclic aromatic and other C10 through C20 unsaturated hydrocarbons. *J. Environ. Sci. Health, Part A: Toxic/Hazard. Subst. Environ. Eng.* **2016**, *51*, 938−953.

(19) Michalík, M.; Lukeš, V. Validation of quantum chemical lipophilicity prediction of alcohols. *Acta Chim. Slovaca* **2016**, *9*, 89−94.

(20) Caballero-García, G.; Mondragón-Solórzano, G.; Torres-Cadena, R.; Díaz-García, M.; Sandoval-Lira, J.; Barroso-Flores, J. Calculation of $V_{S,max}$ and Its Use as a Descriptor for the Theoretical Calculation of pK_a Values for Carboxylic Acids. *Molecules* **2019**, *24*, 79.

(21) Tobiszewski, M.; Tsakovski, S.; Simeonov, V.; Namieśnik, J.; Pena-Pereira, F. *Green Chem.* **2015**, *17*, 4773−4785.

(22) Pena-Pereira, F.; Kloskowski, A.; Namieśnik, J. *Green Chem.* **2015**, *17*, 3687−3705.

(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2016.

(24) Dennington, R.; Keith, T.; Millam, J. *GaussView*, version 5.0; Semichem, Inc.: Shawnee Mission, KS, 2009.

(25) Zhao, Y.; Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(26) Peverati, R.; Truhlar, D. G. Performance of the M11 and M11-L density functionals for calculations of electronic excitation energies by adiabatic time-dependent density functional theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11363−70.

(27) Becke, A. D. Density-Functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−565.

(28) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G Universal solvation model based on solute electron density and a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−96.

(29) Hodges, G.; Eadsforth, C.; Bossuyt, B.; Bouvy, A.; Enrici, M.-H.; Geurts, M.; Kotthoff, M.; Michie, E.; Miller, D.; Müller, J.; Oetter, G.; Roberts, J.; Schowanek, D.; Sun, P.; Venzmer, J. A comparison of log K_{ow} (*n*-octanol−water partition coefficient) values for non-ionic, anionic, cationic and amphoteric surfactants determined using predictions and experimental methods. *Environ. Sci. Eur.* **2019**, *31*, 1.

(30) Sachs, L. *Applied Statistics*; Springer: New York, 1984.