

## QM/NN QSPR Models with Error Estimation: Vapor Pressure and LogP

Bernd Beck,\* Andreas Breindl, and Timothy Clark\*

Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstrasse 25,  
D-91052 Erlangen, Germany

Received October 17, 1999

QSPR models for logP and vapor pressures of organic compounds based on neural net interpretation of descriptors derived from quantum mechanical (semiempirical MO; AM1) calculations are presented. The models are cross-validated by dividing the compound set into several equal portions and training several individual multilayer feedforward neural nets (trained by the back-propagation of errors algorithm), each with a different portion as test set. The results of these nets are combined to give a mean predicted property value and a standard deviation. The performance of two models, for logP and the vapor pressure at room temperature, is analyzed, and the reliability of the predictions is tested.

Estimation of physical properties from molecular structures (*quantitative structure–property relationships, QSPRs*) has been one of the cornerstones of computational chemistry since the pioneering work of Hansch and Leo.<sup>1,2</sup> Many models for properties such as the logarithm of the octanol/water partition coefficient, logP,<sup>3–5</sup> standard enthalpies of formation,<sup>6–8</sup> boiling points,<sup>9</sup> melting points,<sup>10</sup> and aqueous solubility<sup>11,12</sup> are based on incremental approaches in which the molecule is divided into atoms or groups and each fragment is assigned an incremental contribution. With correction factors to take unusual interactions into account, these 2D-methods can be very accurate and are computationally very efficient. They do, however, suffer from the disadvantages that increments may not have been derived for some rare fragments, that exceptional molecular features not found in the parametrization set may render the results unreliable for “exotic” molecules, and that they may tend to be overfitted for some classes of molecules. The typical characteristics of such models is that they fit “normal” molecules very well but that their performance falls very sharply for the “exotics”. This need not necessarily be a serious disadvantage as, for instance, the limitations on candidate drug molecules are relatively restrictive,<sup>13,14</sup> so that property estimation within a limited class of compounds is needed. We<sup>15,16</sup> and others<sup>17,18</sup> have investigated alternative approaches in which quantum mechanically derived 3D-descriptors are used to derive the QSPR, either using linear regression<sup>17</sup> or simple multilayer feedforward neural nets<sup>15,16,18</sup> as the interpretative tool. Our results for logP<sup>15</sup> suggest that models so derived may be very robust and general and that such models may provide a real alternative to the conventional 2D-approaches. 3D-Techniques also offer the possibility that conformationally dependent properties can be calculated. This would, however, require very accurate experimental parametrization data for a large number of conformationally rigid molecules and is presently impractical for almost all physical properties.

Our aim in this work is to derive two quantum mechanics/neural net (QM/NN) QSPR models (for logP and the logarithm of the vapor pressure at 25 °C) using implicit cross-validation of the neural net models and with an estimate of the likely error limits of the predicted values built into the model. We will analyze the performance of the models and investigate a technique for estimating the reliability of the individual predictions.

## DESCRIPTORS

As in our previous work on logP,<sup>15</sup> we use predominantly electrostatic descriptors, including those derived by Politzer et al.,<sup>19</sup> calculated for the AM1<sup>20</sup> optimized structures using the NAO-PC model<sup>21–23</sup> to calculate the molecular electrostatic potential. The descriptors used for the logP nets have been described previously.<sup>15</sup> For the nets used to estimate vapor pressures, the descriptor set defined in Table 1 was used.

**Training/Test Sets.** The set of 1085 molecules used previously for the logP model<sup>15</sup> was also used in this work.

For the parametrization, data for the vapor pressure of compounds measured at 25 °C, or in a temperature range that allowed us to use Antoine’s equation to calculate the vapor pressure at 25 °C, were chosen from ref 24. This results in a total of 551 compounds for the combined training/test set. No data are available that allow us to judge the experimental errors.

**Semiempirical MO-Calculations.** The molecular structures were optimized without symmetry constraints to a gradient norm of 0.4 kcal mol<sup>−1</sup> Å<sup>−1</sup> with VAMP 6.5<sup>25</sup> and VAMP 7.0<sup>26</sup> using the standard default EF optimizer.<sup>27</sup> The standard AM1<sup>20</sup> Hamiltonian and parameter set was used throughout. The optimized geometries for the logP dataset were those defined in ref 15. The starting geometries for the vapor pressure compounds were derived from SMILES strings<sup>28</sup> using CORINA<sup>29</sup> for the 2D-3D conversion. The single conformations resulting from optimization of these starting geometries were used throughout. The molecular electrostatic potentials were calculated using the natural atomic orbital/point charge (NAO-PC) model,<sup>21–23</sup> and

\* Corresponding author phone: ++49 (0)9131 8522948; e-mail: clark@chemie.uni-erlangen.de.

**Table 1.** Descriptor Set for Vapor Pressure Prediction

	property
$\alpha$	mean molecular electronic polarizability
Descriptors Defined by Politzer et al. <sup>19</sup>	
MEP <sub>+</sub>	mean positive electrostatic potential
MEP <sub>-</sub>	mean negative electrostatic potential
V <sub>max</sub>	maximum electrostatic potential
V <sub>min</sub>	minimum electrostatic potential
Sum of the Electrostatic Potential Derived Atomic Charges on	
NSUM	nitrogens
OSUM	oxygens
SSUM	sulfur atoms
HalSUM	all halogen atoms
MW	molecular weight

surface-based descriptors using the solvent-excluded surface (SES) with a solvent radius of 1.4 Å. The SES was calculated using an algorithm based on GEPOL<sup>30</sup> and the Marsili marching cube algorithm.<sup>31</sup> Molecular polarizabilities were calculated using the variational technique of Rivail and Rinaldi.<sup>32</sup> The average computational time needed for the properties calculation on an SGI Power Challenge (195 MHz, R10000) was about 80 s per compound.

**Neural Nets.** Three-layered feedforward neural nets<sup>33–36</sup> were trained using the back-propagation of errors algorithm implemented in Qnet 97<sup>37</sup> and were used under Windows NT 4.0. In the following these neural nets will be denoted “back-propagation neural nets”. Standard sigmoid ( $1/(1 + \exp(-x))$ ) transfer functions were used throughout.

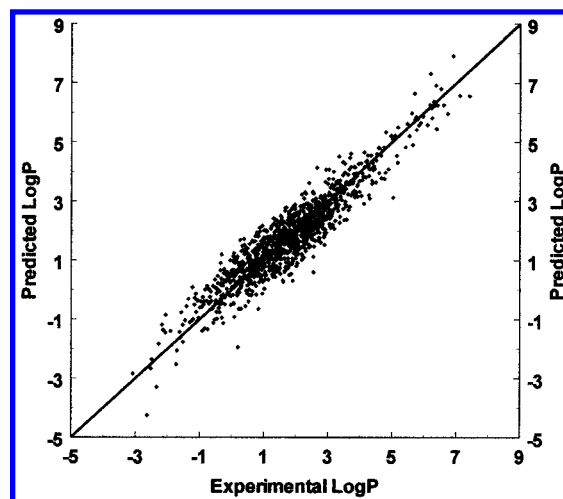
The initial weights for training the back-propagation neural nets were set to random values. Nets were trained to a minimum in the RMS error of the corresponding test set. The predicted power of the trained net was tested by using an external validation set. Training (with different initial weights) was repeated several times for each net/data combination. The results given are from the neural net with the lowest error for the predicted values.

For logP, the dataset was divided into 11 roughly equal randomly selected portions, and each of these portions was used as the test set in turn to train 11 separate nets. For the logarithm of the vapor pressure, the total compound set was divided into 10 nearly equal portions.

**Descriptor and Neural Net Architecture Selection.** Descriptor selection is an iterative procedure. We have used three different techniques for descriptor-preselection. The first is to check the direct correlation of each possible descriptor with the desired property. The second is to perform multiple linear regression analyses and to use the descriptors selected. The final and most promising technique is to perform “Formal Inference-based Recursive Modeling” (FIRM)<sup>38</sup> analyses on the descriptors using the desired property as the lead variable. The descriptors selected to partition the data are well suited as descriptors for the neural nets. In the present study we have used all three methods, also in combination, to obtain the starting descriptor sets. Descriptor sets were then extended according to the nature of the outliers.

Having found several promising sets of descriptors, we start by training several back-propagation neural nets with different numbers of hidden nodes (between 2 and the number of the input nodes).

The best neural nets obtained were analyzed with respect to the weights between the input nodes and the hidden nodes.

**Figure 1.** Cross-validated results for the AM1-NN LogP model (11 nets).

If possible, insignificant descriptors were removed, and new neural nets were trained. If the quality of the neural nets obtained was as good as before we again remove one or more descriptors and train neural nets with different numbers of hidden nodes once again. This procedure is repeated until the predictive power of the neural net declines. Simultaneously we try to enhance the quality of the neural nets by adding or replacing descriptors as described above.

Overall, several hundred back-propagation neural nets were trained in order to find the final net architectures and descriptor sets.

The final network architecture used for logP was 16:10:1 and for the logarithm of the vapor pressure 10:8:1. For the logP and vapor pressure models the resulting training data/connections ratios were 6.38:1 and 6.26:1, respectively.

**Cross-Validation.** The cross-validated result was taken as that given by the nets in which each compound was a member of the test set and thus should represent the worst possible performance of the model for a given compound in the dataset. The prediction of the model for unknown compounds is taken as the simple mean of the predictions of all trained nets, and the error bars are given as one standard deviation based on these 10 or 11 results. In this way, we hope to be able to estimate the likely reliability of the prediction, similar to the OPS technique used by Gombar and Enslein for regression-based models.<sup>39</sup>

## RESULTS AND DISCUSSION

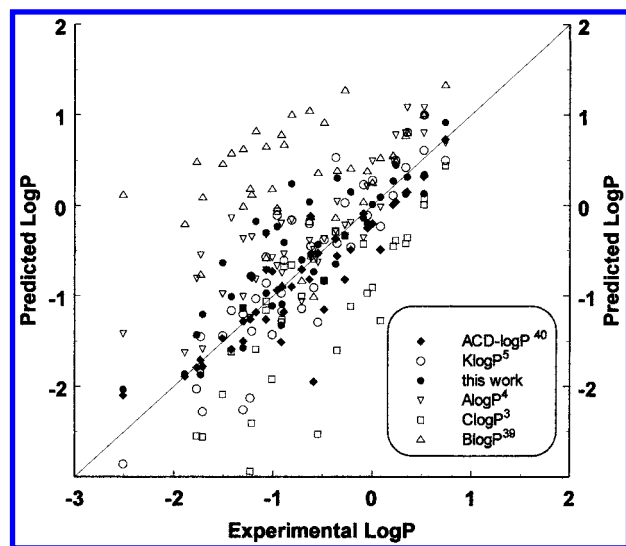
**LogP.** The cross-validation results (16:10:1) obtained for the logP dataset<sup>15</sup> are shown in Figure 1. The standard deviation between predicted and experimental logP values is 0.56, with a maximum error of 2.15. As these results essentially represent the worst net of the 10 for each compound, real performance can be expected to be better.

There is a general, but not universal, tendency for the error limits to be larger at the extremes of the logP range, reflecting the relative paucity of training data in these ranges.

**Comparison to Other Programs.** To test our model with an external dataset and to be able to compare our method with other prediction models we chose a set of 41 nucleosides and nucleoside bases from the literature that were not included in our dataset.<sup>4</sup> Figure 2 shows results for this set.

**Table 2.** Results for Six Different LogP Prediction Methods Applied to the Nucleoside Data Set ( $N = 41$ )

method	linear regression			standard derivation	mean unsigned error	largest error
	$R^2$	constant	slope			
AM1/NN	0.74	$0.22 \pm 0.08$	$0.92 \pm 0.09$	0.39	0.32	-1.05
ACD <sup>41</sup>	0.83	$0.09 \pm 0.08$	$0.97 \pm 0.07$	0.32	0.22	1.36
KlogP <sup>5</sup>	0.81	$-0.22 \pm 0.06$	$0.68 \pm 0.05$	0.34	0.36	1.20
AlogP <sup>4</sup>	0.74	$-0.39 \pm 0.07$	$0.95 \pm 0.09$	0.39	0.42	-1.28
ClogP <sup>3</sup>	0.59	$0.01 \pm 0.12$	$0.56 \pm 0.08$	0.34	0.65	2.04
BlogP <sup>40</sup>	0.16	$-0.86 \pm 0.13$	$0.52 \pm 0.19$	0.70	1.05	-2.62

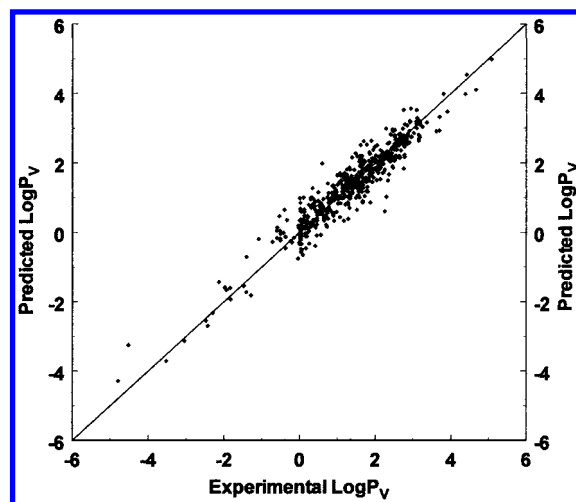
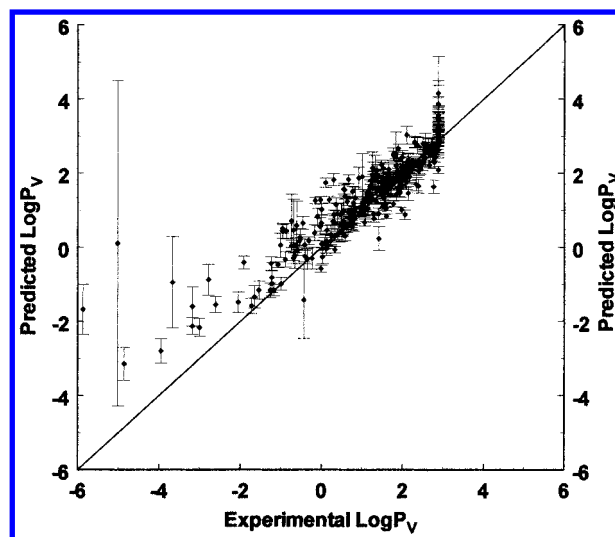
**Figure 2.** LogP values calculated using six different methods for a set of 41 nucleosides and nucleoside bases.<sup>3,4,5,15,40,41</sup>

The corresponding values for five other available estimation methods<sup>3-5,40,41</sup> are also shown in the plot. The geometries were generated using a standardized procedure in which the 2D-structure is converted to 3D with CORINA,<sup>29</sup> and the resulting 3D structure is optimized using Cartesian coordinates with VAMP 7.0.<sup>26</sup> This procedure was used<sup>42</sup> to convert the entire Maybridge database<sup>43</sup> and represents an attempt to derive a consistent single-conformation dataset.

For this set of unknowns, the standard deviation is 0.39, and the maximum absolute error is 1.05. Table 2 shows a summary of the performance of all six estimation methods for this dataset. As shown in Table 2 the ACD-LogP<sup>41</sup> and KlogP<sup>5</sup> fragment based methods perform best with respect to the standard deviation between calculated and experimental values. However, the largest absolute error is lowest for our model, confirming the robustness of the QM/NN technique. We note here that compounds of the type found in this dataset were excluded specifically from the original parametrization set for the QM/NN technique.<sup>15</sup>

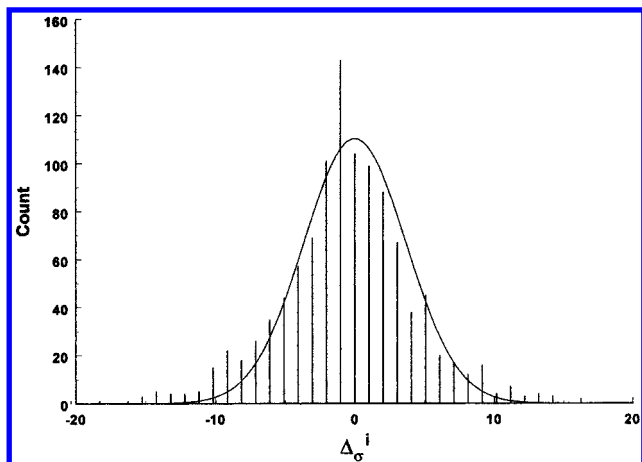
**Vapor Pressure.** Figure 3 shows the cross-validated result for the vapor pressure dataset. The cross-validation gives a standard deviation of 0.37 log units (Torr) and a maximum absolute error of 1.65.

Figure 4 shows the results obtained for a set of 192 unknown compounds from the Registry of Physicochemical Data.<sup>44</sup> The error bars given in Figure 4 were derived using the procedure outlined below and given in eq 2. For this dataset, the standard deviation is 0.68, and the two largest errors are 5.12 and 4.19. These extreme outliers occur because our original dataset contains no phosphorus compounds, and no experimental values below -4.8. 110 (57%) of the results agree within one standard deviation.

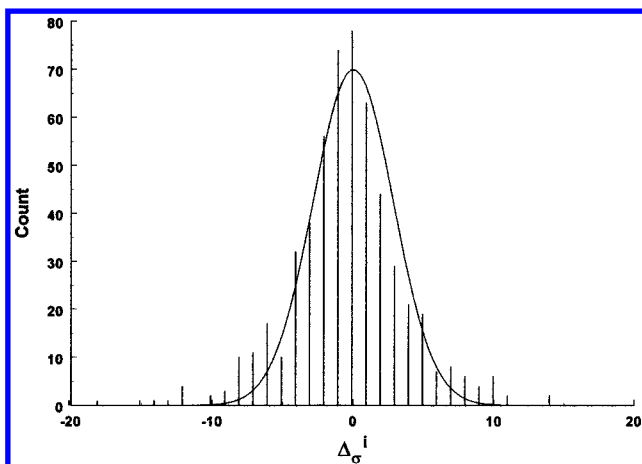
**Figure 3.** Cross-validated results for the AM1-NN vapor pressure model (10 nets).**Figure 4.** Results for the vapor pressures of 192 compounds not included in the training/test dataset and taken from the registry of physicochemical data.<sup>44</sup>

Gallagher et al.<sup>45</sup> have published a method for predicting vapor pressure by using the calculated molecular polarizability and the counts of six different polar groups within a regression model. For their dataset of 479 compounds they obtain a correlation coefficient ( $r^2$ ) of 0.96 with a standard deviation of 0.534. A neural net was also tested as the interpolation technique but gave no major improvement.

**Error Estimation for Individual Compounds.** The procedure used above equates the spread of values estimated by the different nets with the likely reliability of the prediction. To gain some insight into the validity of this assumption, we have plotted histograms of the deviation between calculated and experimental values for compound



**Figure 5.** The distribution of  $\Delta\sigma^i$ -values for the 11-net logP model. The solid line shows the best fit Gaussian.



**Figure 6.** The distribution of  $\Delta\sigma^i$ -values for the 10-net vapor pressure model. The solid line shows the best fit Gaussian.

$i$ , for  $\Delta\sigma^i$  expressed in terms of the predicted individual standard deviation for that compound,  $\sigma^i$

$$\Delta\sigma^i = \frac{(P_{calc}^i - P_{obs}^i)}{\sigma^i} \quad (1)$$

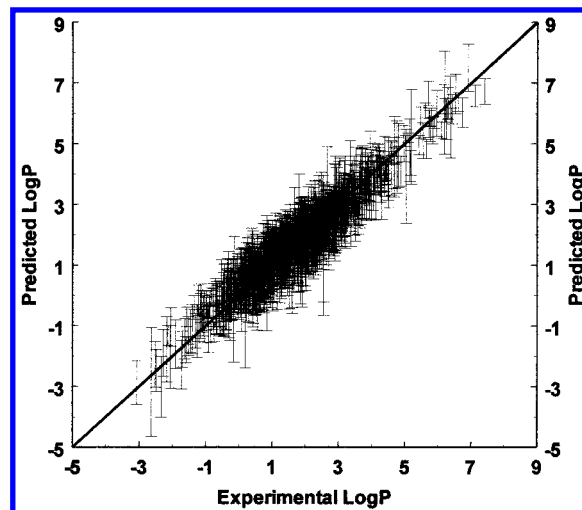
where  $\sigma^i$  is given by

$$\sigma_i = \frac{1}{N-1} \sqrt{\sum_{j=1,N} (X_i - P_j^i)^2} \quad (2)$$

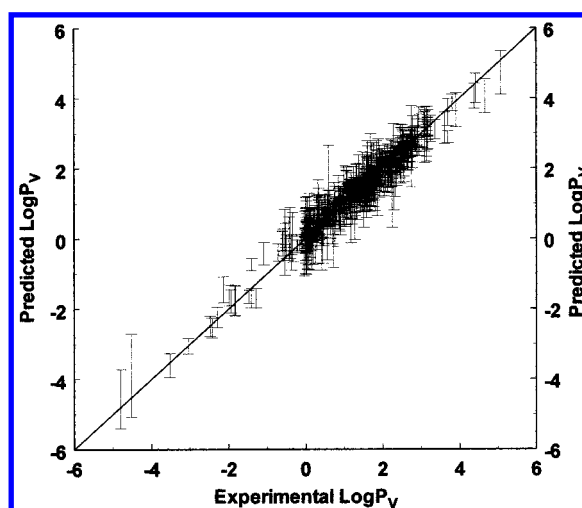
where  $N$  is the number of individual neural nets,  $X_i$  is the experimental value for compound  $i$ , and  $P_j^i$  is the predicted value for compound  $i$  by net  $j$ .

The histogram for the logP training/test dataset is shown in Figure 5, and the corresponding plot for log(vapor pressure) is shown in Figure 6. The distribution of the errors expressed in units of the predicted individual standard deviation are compared with best fit Gaussian curves. These curves are also shown in Figures 5 and 6.

The results show that the error distributions of the two prediction models are very similar to Gaussian distributions with half widths of 8.54 and 6.81 and mean absolute values of  $\Delta\sigma^i$  of 3.58 and 2.98 for logP and vapor pressure models, respectively. An analysis of the unsigned errors for the two models gives mean values and standard deviations of 0.38



**Figure 7.** Results for the logP dataset ( $N = 1085$ ) with the mean values and error estimations given in eq 2.



**Figure 8.** Results for the vapor pressure dataset ( $N = 551$ ) with the mean values and error estimations given in eq 2.

$\pm 0.28$  and  $0.22 \pm 0.19$ , respectively, compared with  $0.14 \pm 0.11$  and  $0.10 \pm 0.06$  for the standard deviations of the net predictions. We therefore suggest that realistic estimates of the probably reliability of the net predictions can be obtained by multiplying the standard deviation of the net predictions by the mean of the  $\Delta\sigma^i$ -values for the training dataset. This procedure would lead to error estimates for the logP and vapor pressure datasets of  $0.50 \pm 0.39$  and  $0.30 \pm 0.18$ , respectively, slightly larger than the observed values. Thus, the property value predicted by  $N$  neural nets is given as

$$P^i = \frac{\sum_{j=1,N} P_j^i}{N} \pm \bar{\Delta}\sigma^i \quad (3)$$

where  $P^i$  is the predicted property for compound  $i$  and  $\bar{\Delta}$  is the mean of the  $\Delta\sigma^i$ -values for the training set. Figures 7 and 8 show error bar plots derived using eq 2 for the logP and vapor pressure models, respectively. The error bars estimated from eq 2 are also shown in Figure 4. These three plots illustrate the usefulness of this error estimation technique. The vapor pressure model, for instance, suffers from a lack of reliable training data at the lower (negative)



end of the scale and tends to give larger error bars in this region. For the logP and vapor pressure training sets (Figures 7 and 8, respectively), 63% and 64% of the data are within the estimated error limits. Thus, the error bars estimated by this procedure correspond roughly to the confidence limits given by  $\pm$  one standard deviation. Note that the above procedure is purely empirical in that our experience and the tests described above indicate that it gives useful results. A leave-one-out cross-validation would not serve our purposes here as it would not give an error estimate for individual compounds, which is our aim. In this respect it is more akin to an OPS analysis,<sup>39</sup> rather than more general tests of model reliability in general, such as cross-validation or confidence intervals.

## CONCLUSIONS

The techniques presented here provide general and robust models for the estimation of physical properties that are not dependent on the availability of data to derive individual increments. The validation of the models by the multinet technique provides a useful and informative assessment of the likely reliability of the model itself, and the error estimates for individual compounds help to identify compounds for which the model is less reliable. However, the limiting factor in the development of such models remains the availability of sufficient consistent and reliable data. One further important aspect of such models is their conformational dependence, which will be investigated in further work.<sup>46</sup>

## ACKNOWLEDGMENT

We thank Oxford Molecular for financial support and John Holland and Andrew Chalk for useful discussions.

## REFERENCES AND NOTES

- (1) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients And Their Uses. *Chem. Rev.* **1971**, 71, 524–616.
- (2) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- (3) Leo, A. J. *ClogP*; Daylight Chemical Information Systems: Irvine, CA, 1991.
- (4) Viswanadhan, V. N.; Reddy, M. R.; Bacquet, R. J.; Erion, D. M. Assessment of Methods Used for Predicting Lipophilicity: Application to Nucleosides and Nucleoside Bases. *J. Comput. Chem.* **1993**, 9, 1019–1026.
- (5) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer Automated logP Calculations Based on an Extended Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 752–781.
- (6) Benson, S. W. *Thermochemical Kinetics*, 2nd ed.; Wiley: New York, 1976.
- (7) Clark, T.; McKervey, M. A. Saturated Hydrocarbons. In *Comprehensive Organic Chemistry*; Barton, D. H. R., Ollis, W. D., Eds.; Pergamon Press: Oxford, 1979; Vol. 1, Chapter 2, pp 37–120.
- (8) Cohen, N.; Benson, S. W. In *Chemistry of alkanes and cycloalkanes*; Patai, S., Rappoport, Z., Eds; Wiley: Chichester, 1992; Chapter 6, p 215.
- (9) Stein, S. E.; Brown, R. L. Estimation of normal boiling points from group contributions. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 581–587.
- (10) Constantinou, L.; Gani, R. New group contribution method for estimating properties of pure compounds. *AIChE J.* **1994**, 40, 237–244.
- (11) Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of aqueous solubility of organic molecules by the group contribution approach. Application to the study of biodegradation. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 474–482.
- (12) Kuhne, R.; Ebert, R.-U.; Kleint, F.; Schmidt, G.; Schuurmann, G. Group contribution methods to estimate water solubility of organic chemicals. *Chemosphere* **1995**, 30, 2061–2077.
- (13) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Rev.* **1997**, 23, 3–25.
- (14) Sadowski, J.; Kubinyi, H. A Scoring Scheme for Discriminating between Drugs and Nondrugs. *J. Med. Chem.* **1998**, 41, 3325–3329.
- (15) Breindl, A.; Beck, B.; Clark, T.; Glen, R. C.; Prediction of the *n*-Octanol/Water Partition Coefficient, logP, Using a Combination of Semiempirical MO-Calculations and a Neural Network. *J. Mol. Model.* **1997**, 3, 142–155.
- (16) Clark, T.; Breindl, A.; Rauhut, G. A Combined Semiempirical MO/Neural Net Technique for Estimating <sup>13</sup>C Chemical Shifts. *J. Mol. Model.* **1995**, 1, 22–35.
- (17) Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, M. E. Politzer, P. Relationships of Critical Constants and Boiling Points to Computed Molecular Surface Properties. *J. Phys. Chem.* **1993**, 97, 9369–9373.
- (18) Parham, M.; Hall, L. Accurate Prediction of *n*-Octanol/Water Partition Coefficient Using Neural Network Algorithms and E-State Atom Indices; IBD Drug Discovery Conference, Boston, August 1999.
- (19) Murray, J. S.; Politzer, P. Statistical analysis of the molecular surface electrostatic potential: an approach to describing noncovalent interactions in condensed phases. *J. Mol. Struct. (THEOCHEM)* **1998**, 425, 107–114.
- (20) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, 107, 3902–3909.
- (21) Rauhut, G.; Clark, T. Multicenter Point Charge Model for High Quality Molecular Electrostatic Potentials from AM1 Calculations. *J. Comput. Chem.* **1993**, 14, 503–509.
- (22) Beck, B.; Rauhut, G.; Clark, T. The Natural Atomic Orbital Point Charge Model for PM3: Multipole Moments and Molecular Electrostatic Potentials. *J. Comput. Chem.* **1995**, 15, 1064–1073.
- (23) Beck, B.; Clark, T. Some Biological Applications of Semiempirical MO Theory. *Perspectives Drug Discovery Design* **1998**, 9/10/11, 131–159.
- (24) Dean, J. A. *Lange's Handbook of Chemistry*, 12th ed.; McGraw-Hill Book Co.: New York, 1979.
- (25) Rauhut, G.; Alex, A.; Chandrasekhar, J.; Steinke, T.; Sauer, W.; Beck, B.; Hutter, M.; Gedeck, P.; Clark, T. VAMP 6.5; Oxford Molecular Ltd.: Medawar Centre, Oxford Science Park, Sandford-on-Thames, Oxford, OX4 4GA, U.K., 1997.
- (26) Clark, T.; Alex, A.; Beck, B.; Chandrasekhar, J.; Gedeck, P.; Horn, A.; Hutter, M.; Rauhut, G.; Sauer, W.; Steinke, T. VAMP 7.0; Oxford Molecular Ltd.: Medawar Centre, Oxford Science Park, Sandford-on-Thames, Oxford, OX4 4GA, U.K., 1998.
- (27) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1985**, 7, 385–395.
- (28) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31–36.
- (29) Sadowski, J.; Gasteiger, J. *Corina v. 1.8*; Oxford Molecular: Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, UK.
- (30) Pascual-Ahuir, J. L.; Silla, E.; Tuñon, I. GEOPOL: An Improved Description of Molecular Surfaces. III. A New Algorithm for the Computation of a Solvent-Excluding Surface. *J. Comput. Chem.* **1994**, 15, 1127–1138.
- (31) Marsili, M. In *Physical Property Prediction in Organic Chemistry*; Jochum, C., Hicks, M. G., Sunkel, J., Eds., Springer: Berlin-Heidelberg, 1988; p 249.
- (32) Rinaldi, D.; Rivail, J. Calculation of molecular electronic polarisabilities. Comparison of different methods. *Theor. Chim. Acta* **1974a**, 32, 243–251.
- (33) Pao, Y.-H. *Adaptive Pattern Recognition and Neural Networks*; Addison-Wesley Publishing Company: Reading, 1989.
- (34) Müller, B.; Reinhardt, J.; Strickland, M. T. *Neural Networks – An Introduction*, 2nd ed.; Springer-Verlag: Berlin, Heidelberg, 1995.
- (35) Rojas, R. *Neural Networks, A Systematic Introduction*; Springer-Verlag: Berlin, Heidelberg, 1996.
- (36) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH Verlag: Weinheim, 1993.
- (37) Qnet97; Vesta Service Inc.: 1001 Grenn Bay Rd., Box 196, Winnetka, IL 60093, USA. (b) Back-propagation Technical Overview at <http://www.qnetv2k.com/Qnet2000Manual/contents2000.htm>.
- (38) Hawkins, D. M. *FIRM*; <http://www.stat.umn.edu/users/FIRM/index.html>.
- (39) Gombar, V. K.; Enslein, K. Assessment of *n*-octanol/water partition coefficient: when is the assessment reliable? *J. Chem. Inf. Comput. Sci.* **1996**, 36, 1127–1134.
- (40) Bodor, N.; Gabanyi, Z.; Wong, C.-K. A New Method for the Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, 111, 3783–3786.
- (41) Viswanadhan, V. N.; Reddy, M. R.; Bacquet, R. J.; Erion, M. D. *J. Comput. Chem.* **1993**, 14, 1019–1026. (b) Klopman, G.; Li, J.-Y.;

- Wang, S.; Dimayuga M. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 752–781. (c) [http://www.acdlabs.com/products/phys\\_chem\\_lab/logp/competit.html#Table 1](http://www.acdlabs.com/products/phys_chem_lab/logp/competit.html#Table 1).
- (42) Beck, B.; Horn, A.; Carpenter, J. E.; Clark, T. Enhanced 3D-Databases: A Fully Electrostatic Database of AM1-Optimized Structures. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 1214–1217.
- (43) Maybridge Chemicals Company Ltd., Trevillet, Tintagel, Cornwall PL34 OHW, England.
- (44) Baricic, P.; Mackov, M. *The Registry of Physicochemical Data, Version 1.2.2*; Synexchem Consulting Services International, LLC: 1999.
- (45) Liang, C.; Gallagher, D. A. QSPR Prediction of Vapor Pressure from Solely Theoretically-Derived Descriptors. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 321–324.
- (46) Chalk, A., Beck, B., Clark, T., manuscript in preparation.

CI990131N