Molecular Parameters Responsible for the Melting Point of 1,2,3-Diazaborine Compounds

Boris Johnson-Restrepo, Leonardo Pacheco-Londoño, and Jesus Olivero-Verbel*

Environmental and Computational Chemistry Group, Department of Chemistry, University of Cartagena,
Cartagena, Colombia, South America

Received December 4, 2002

Quantitative structure−property relationships (QSPR) have been determined to predict the melting point temperatures of 1,2,3-diazaborine compounds ($n = 72$). Electronic and topological descriptors were computed from molecular structures, and a QSPR model was generated by linear multiple regression using reported melting point temperatures as the dependent variable. The most important molecular descriptors describing this physicochemical property were the sum of the atomic charges for the heteroatoms, the sum of Randić connectivity indexes $^0X^0$, $^0X^1$, and $^0X^2$, the total number of atoms in the molecule, and the volume of the box in which the molecule fits. The multiple determination coefficient (R^2) and standard error of estimation for the model were 0.856 and 16.787 °C, respectively. In addition to regression techniques, a back-propagation neural network was used to include nonlinear relationships between molecular structure and melting point temperatures. It is concluded that melting point temperatures for 1,2,3-diazaborine compounds can be described by electrostatic interactions mediated by atomic charges and steric properties. The results of this study demonstrate that multiple linear regression analysis and back-propagation neural network are techniques that can be used to successfully predict the melting point temperatures of 1,2,3-diazaborine compounds. The most accurate prediction results were obtained using the Levenberg−Marquardt neural network algorithm.

INTRODUCTION

1,2,3-Diazaborines are a family of compounds with antibacterial properties against a number of Gram-negative bacteria and other organisms.[1,2] These agents block lipid biosynthesis by specifically inhibiting the enzyme enoyl-acyl carrier protein reductase,[3,4] reducing the enoyl group of a fatty acid chain attached to the acyl carrier protein, to its acyl product using NAD(P)H as the cofactor.[5] In addition, it has been shown that diazaborines can stabilize aberrant mRNAs in yeasts.[6] 1,2-Dihidro-1-hydroxy-2-(organosulfonyl)-areno[d][1,2,3]diazaborines (areno = benzene, naphthalene, thiophene, furan, pyrrole) can be synthesized by reaction of (organosulfonyl) hydrazones of arene aldehydes or ketones with tribromoborane in the presence of $FeCl_3$.[7] A more general synthesis squeme for diazaborines has been recently reported and includes the preparation of amidines from anilines using a nitrile and $AlCl_3$, followed by reaction with BCl_3.[8]

The study of the physicochemical and biological properties of these organic compounds, and the molecular features responsible for them, is fundamental in the process of designing new therapeutic agents based on the diazaborine skeleton.[7] An important tool necessary for this process is the development of quantitative structure−property relationship (QSPR) models. These are statistical equations that correlate 2D-(two-dimension) or 3D-(three-dimension) descriptors obtained from molecular structure and any kind of physicochemical or biological properties associated with them. Using this approach, it is possible to generate functions that are able to predict molecular properties for compounds without experimental measurements. In general 2D-based descriptors are more useful in predicting some physicochemical properties such as logP and pK_a, compared to 3D descriptors.[9−11] 3D-QSPR methods have been used as procedures to explain the variance in physicochemical or biological properties by monitoring variations in the 3D structure of the chemical compounds.[9] CoMFA[12] (Comparative Molecular Field Analysis), GRID,[13] and PLS[14] (partial least squares) are the best methods in 3D-QSPR. In the case of melting point (MP) temperatures of 1,2,3-diazaborines, molecules with fused planar rings containing few substitutions, the relevance of conformational flexibility that could be encoded in 3D-QSPR is low for a property based in a solid state, in contrast with the majority of biological properties. However, MP temperature can help us to understand biochemical properties such as toxicity[15,16] and other physicochemical variables such as reticular energy and water solubility.[17] Besides its importance, few reports have been published about QSPR models for MP temperatures[18−21]

Based on the relevance of 1,2,3-diazaborines as new therapeutic agents, whose structure−activity relationships have been previously established,[22] and the diversity in MP temperatures for these compounds, the objective of this work was to develop a QSPR model in order to decode those molecular features responsible for this physicochemical property.

Together with QSAR models obtained by statistical analysis, neural network methods have also been successfully used to consider nonlinear associations between molecular structure and properties, giving the opportunity to obtain

* Corresponding author phone: 57-56698179; fax: 57-56698323; e-mail: jesusolivero@yahoo.com. Corresponding author address: A.A. 6541, Cartagena, Colombia, South America.

10.1021/ci020387k CCC: $25.00 © 2003 American Chemical Society
Published on Web 08/30/2003

much better prediction results when including molecules that were not in the data set used to built the model. There are many applications in structure−activity relationships that use neural network algorithms such as back-propagation (BP)[23−33] or radial function.[34−35] In this paper BP neural network and some of its variations were used to get faster convergence speed. Excellent tutorial and review articles about BP algorithm can be found elsewhere.[25,26,36,37]

The BP is the best known training algorithm for neural networks and still one of the most useful for QSAR research. A typically BP neural network consists of three layers fully connected and feed forward. The first layer is known as the input layer and contains a number of neurons equal to the number of the molecular descriptors obtained from the best multiple linear regression analysis used to define the QSAR. This layer transmits a signal to the nodes of the second layer or hidden layer, which processes data using a sigmoid transfer function, and sends them to the output layer that contains one neuron representing the predicted output, also calculated with a sigmoid function. A neural network "learns" by passing repeatedly data through the neurons and adjusting their weights and biases to minimize the Mean Squared Error (MSE), $\sum(t_i-o_i)^2/n$ (i is an index for training observations and n is the total number of input vectors), until the network predicted output (o_i) matches the target (t_i) or given property point values. In BP neural networks, the updated weights and biases are obtained using a gradient descent method, in which the partial derivative of the error function is used to determine each weight adjustment, (ΔW_{ij}), as follows

$$\Delta W_{i,j}(n) = \eta\delta_i O_j + \alpha\Delta W_{i,j}(n-1) \qquad (1)$$

where ΔW_{ij}, is the change in the weight factor for each network node i in the layer j, δ_i is the actual error of node i, and O_j is the output of the node j. The coefficients η and α are the learning rate and the momentum factor, respectively.

A neural network can be trained to perform a particular task using a variety of learning algorithms. In the case of BP, there are several variations on the standard algorithm which are based on other optimization techniques, that allow better speed convergence and fitting models. These methods include the Levenberg−Marquardt, quasi-Newton, and conjugate gradient algorithms.

METHODS

The process of QSPR model generation includes the selection of an experimental data set, computer-assisted molecular modeling of the compounds to obtain molecular descriptors, statistical analysis to develop the model, and finally a validation process.[38,39]

Experimental Data Set. Seventy-two MP temperatures for 1,2,3-diazaborines were taken from the literature.[7] Molecular structure for all 1,2,3-diazaborine families and individual compounds used in the study are shown in Figure 1 and Table 1, respectively. MP temperatures for the 1,2,3-diazaborines included in the data set oscillated between 60 and 262 °C, with an average value of 171.3 °C. In general, diazaborines used in this study had a common six-membered heterocyclic ring in the structure, and the differences between them mainly were raised from the substitutions in N1 and C2−C3.

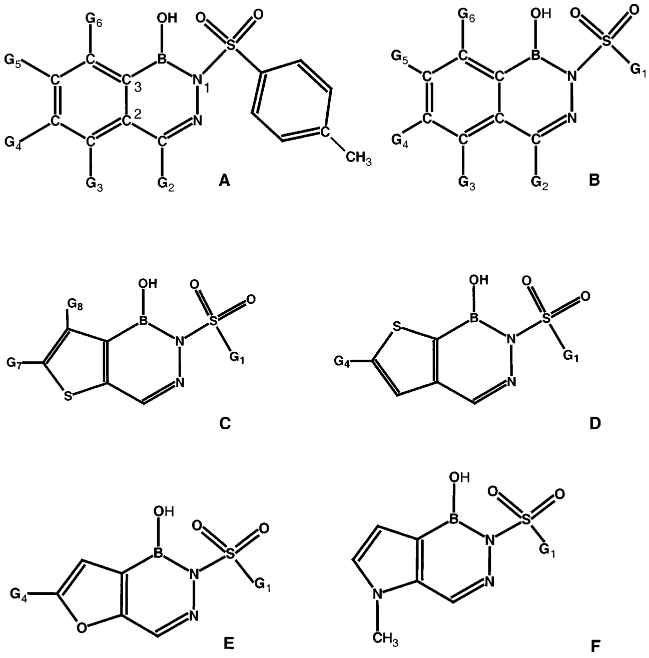


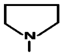
Figure 1. Molecular structure of different 1,2,3-diazaborine analogues included in the study.

Molecular Descriptor Calculation. A total number of 67 topological and electronic descriptors were calculated. Quantum molecular calculations for electronic descriptors from the chemical structure of the 1,2,3-diazaborines were performed through computer calculations on a Digital Alpha Work Station. Initial coordinates were generated after full AM1 optimization with the Spartan program,[40] and single point data were calculated through ab initio molecular quantum mechanics (RHF/6-31G(d)) method using the Gaussian 94 program.[41] Quantum chemically derived descriptors included total energy, energies of frontier orbitals, dipolar moment, quadrupolar moments, and atomic charges, among others. Different combinations between molecular parameters as well as their numerical transformations (inverse, square, root square, or logarithm) were also used as additional descriptors. Geometrical descriptors such as moments of inertia I, II, and III as well as box volume were calculated from optimized structures. Topological descriptors were obtained using graph-theoretical methods to capture molecular information. These parameters were generated using PCDM,[42] a program written in turbo Pascal by our research group. Calculated descriptors included Randić, Hall and connectivity indexes and fragment descriptors such as number of different atoms in the molecule.[43,44]

QSPR Model. The first step to develop the prediction model was to select a group of molecular descriptors encoding particular electronic, topological, or geometrical information about the 1,2,3-diazaborines. Initially, pairs of descriptors were correlated, and those with correlation coefficients greater than 0.9 were additionally correlated with the physicochemical property, the MP temperature. The descriptor with the best correlation was then used for further analysis, leaving out the descriptor having the lower correlation.[45] The relationship between MP temperatures and the molecular descriptors was performed using forward stepwise[46] Multiple Linear Regression (MLR) to generate an equation of the form MP $= a_1 + a_2X_2 \ldots a_nX_n$, where $X^2\ldots,X_n$ and $a_1,a_2\ldots a_n$ are the descriptors and the regression

Table 1. Molecular Structures of 1,2,3-Diazaborines in the Data Set

GROUP A

Molecule	G2	G3	G4	G5	G6
dzb01	H	H	H	H	H
dzb02	CH_3	H	H	H	H
dzb03	H	CH_3	H	H	H
dzb04	H	H	CH_3	H	H
dzb05	H	CH_3	H	H	CH_3
dzb06	H	F	H	H	H
dzb07	H	H	F	H	H
dzb08	H	H	H	F	H
dzb09	H	Cl	H	H	H
dzb10	H	H	Cl	H	H
dzb11	H	H	H	Cl	H
dzb12	H	Cl	H	Cl	H
dzb13	H	H	Cl	Cl	H
dzb14	H	Br	H	H	H
dzb15	H	H	Br	H	H
dzb16	H	H	H	Br	H
dzb17	H	H	H	OH	H
dzb18	H	H	NH_2	H	H
dzb19	H	H	$N(CH_3)$	H	H
dzb20	H	H	H	$N(CH_3)_2$	H
dzb21	H	H	H	$NHCOCH_3$	H
dzb22	H	H		H	H
dzb23	H	Cl	H	$N(CH_3)_2$	H
dzb24	H	H	H	COOH	H
dzb25	C_6H_5	H	H	H	H

GROUP B

	G1	G2	G3	G4	G5	G6
dzb26	C_6H_5	H	H	H	OH	H
dzb27	2,4,6-$(CH_3)_3C_6H_2$	H	F	H	H	H
dzb28	2,4,5-$Cl_3C_6H_2$	H	H	Br	H	H
dzb29	4-$H_2NC_6H_4$	H	F	H	H	H
dzb30	4-$H_2NC_6H_4$	H	H	Br	H	H
dzb31	4-$H_2NC_6H_4$	H	H	CH_3	H	H
dzb32	2-Cl-4-$H_2NC_6H_3$	H	H	CH_3	H	H
dzb33	2-Cl-4-$CH_3CONHC_6H_3$	H	H	CH_3	H	H
dzb34	2-Cl-4-$CH_3CONHC_6H_3$	H	H	Br	H	H
dzb35	4-$O_2NC_2H_4$	H	H	Br	H	H
dzb36	CH_3	H	H	H	H	H
dzb37	CH_3	H	H	CH_3	H	H
dzb38	n-C_3H_7	H	H	CH_3	H	H
dzb39	n-C_3H_7	H	H	Cl	H	H
dzb40	$(CH_3)_2N$	H	H	H	H	H
dzb41	$(CH_3)_2N$	H	H	CH_3	H	H

GROUP C

	G1	G7	G8
dzb42	4-$CH_3C_6H_4$	Br	H
dzb43	4-$CH_3C_6H_4$	H	Br
dzb44	C_6H_5	Br	H
dzb45	2-$CH_3C_6H_4$	Br	H
dzb46	2-$CH_3C_6H_4$	CH_3	H
dzb47	2-ClC_6H_4	Br	H
dzb48	2-ClC_6H_4	CH_3	H
dzb49	2-ClC_6H_4	C_2H_5	H
dzb50	2-Cl-4-CH3C6H3	Br	H
dzb51	2-Cl-4-CH3C6H3	CH_3	H
dzb52	4-$CH_3C_6H_4$	Cl	H
dzb53	2,4,6-$(CH_3)_3C_6H_4$	Br	H
dzb54	4-$CH_3CONHC_6H_3$	Br	H
dzb55	2-Cl-4-CH3CONHC_6H_3	Br	H
dzb56	CH_3	H	H
dzb57	C_2H_5	Br	H
dzb58	n-C_3H_7	CH_3	H
dzb59	$(CH_3)_2CHCH_2$	Br	H

GROUP D

	G1	G4
dzb60	4-$CH_3C_6H_4$	Br
dzb61	4-$CH_3C_6H_4$	C_2H_5
dzb62	2-ClC_6H_4	Br
dzb63	2-ClC_6H_4	C_2H_5
dzb64	2-Cl-4-CH3C6H3	CH_3
dzb65	n-C_3H_7	CH_3
dzb66	n-C_3H_7	C_2H_5
dzb67	$(CH_3)_2CHCH_2$	CH_3

GROUP E

	G1	G4
dzb68	4-$CH_3C_6H_4$	CH_3
dzb69	4-$CH_3C_6H_4$	Br
dzb70	2,4,5-$Cl_3C_6H_2$	Br

GROUP F

	G1
dzb71	C_3H_7
dzb72	4-$CH_3C_6H_4$

coefficient parameters, respectively. The quality of the regression model was measured using primarily four statistical parameters: the determination coefficient (R^2), standard error of estimation (SE), F-value from the analysis of variance (F), and the significance level value (P).[47] If the R^2-value is close to 1, this indicates that the model is self-consistent and can be used to predict the MP temperatures of other 1,2,3-diazaborines within the range of MP temperatures in the data set. The robustness of the obtained model was evaluated using internal jackknifing validation.[48] In this method, each molecule was successively removed from the data set, and its MP temperature was calculated from a new model built from the remaining compounds. This procedure was done for each one of the molecules in the data set, and the predicted values were then correlated with the experimental observations. The generated correlation coefficient is called the cross-validation coefficient (R_{Cross}),[38] and a high quality model is derived when this value is close to one. Moreover, variance inflation factors (VIF) were calculated in order to detect multicollinearities in the model using the formula VIF $= 1/(1-R^2)$. For each selected descriptor the R^2 value was calculated by regressing each descriptor against all others without including the dependent variable.[49] A VIF lower than 10 indicates that the model has been originated from independent molecular properties.

Neural Network. A BP neural network was used in this study, which consisted of a fully connected three-layer system. Each neuron in a given layer was fully connected to all neurons in the adjacent levels. The number of neurons in the input layer was equal to the number of the molecular descriptors taken from the best MLR analysis. The number of neurons in the hidden layer was optimized by trial and error assays calculating the MSE of the training and cross-validation sets in the training process. The output layer contained one neuron representing the predicted MP temperature values. The network was trained using the standard BP algorithm, implementing some variations such as the quasi-Newton and the Levenberg−Marquardt, incorporated as built functions in the MATLAB Neural Network Toolbox 3.0,[36] installed in a PC computer. A hyperbolic tangent sigmoid transfer function, $\tanh(x) = (\exp(x)-\exp(-x))/(\exp(x)+\exp(-x)))$, was used to calculate the output of the neuron in the hidden layer. The log sigmoid transfer function, $\mathrm{logsig}(x) = 1/(1 + \exp(-x))$, was employed to calculate the output in the neuron belonging to the output layer, giving restricted values to the interval between 0 and 1. Therefore, the output values were transformed from the interval 0−1 to the actual MP temperatures. The values that feed the input layer (descriptor values) to the neural network process were normalized so that they had mean values of zero and standard deviations of 1. Similarly, the initial weighs and biases were generated randomly. Both the learning rate and the momentum were optimized by trial and error and the best value obtained for these parameters were 0.0001 and 0.01, for the learning rate and momentum, respectively.

To train the neural network, the data were randomly divided into two groups, a training set and a validation set consisting of 65 and 7 compounds, respectively. This split process was repeated five times to acquire enough data to compute the cross validation coefficient.[25,19] The training set used to update the network weights and biases, and the validation set were also employed to monitor the error during the training process. The training procedure for each random division was repeated 10 times with different initial weights and biases, yielding a total of 50 training experiments. The training set was used for the model generation, and the validation sets were employed to monitor the training process and to evaluate the generated model.

Table 2. Four-Descriptor Multilinear Regression Model Selected for Compounds in Data Set[a]

descriptor	coefficient	standard error	T-statistic	*p*-value	VIF
ΣAC_{HET}	−13.35	4.0	−3.32	0.0014	1.6
Σ^0X	17.50	1.4	12.64	<0.0001	9.8
NA	−10.40	1.0	−10.48	<0.0001	4.7
V_{BOX}	−0.0046	0.0012	−3.68	0.0005	4.7
intercept	−39.98	17.5	−2.29	0.0014	

[a] Model statistics: $R^2 = 0.856$; SE = 16.787; *P*-value <0.0001; $R^2_{Cross} = 0.834$; *F*-ratio = 99.250, $N = 72$.

Table 3. Individual Correlation Coefficient between Molecular Descriptors Used in the Model and the Respective Melting Point Temperatures for All the Molecules in the Data Set

descriptor	correlation coefficient (*R*)
Σ^0X	0.76
NA	0.37
ΣAC_{HET}	−0.42
V_{BOX}	0.58

Table 4. One- to Four-Descriptor Stepwise Multiregression Model for Melting Point Temperatures of 72 Diazaborines[a]

descriptor	R^2	*S*	R^2_{Cross}	S_{Cross}
Σ^0X	0.582	27.958	0.542	29.239
Σ^0X + NA	0.776	20.626	0.744	21.885
Σ^0X + NA + V_{BOX}	0.832	17.973	0.812	18.755
Σ^0X + NA + V_{BOX} + ΣAC_{HET}	0.856	16.787	0.834	17.683

[a] R^2 determination coefficient; *S* standard error of estimate; R^2_{Cross} leave-one-out cross-validated determination coefficient; S_{Cross} leave-one-out cross-validated standard error of the estimate.

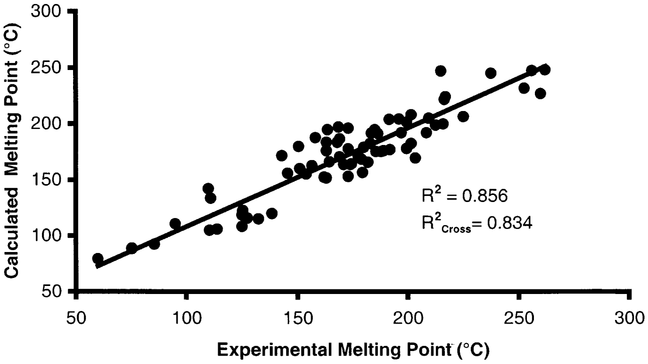


Figure 2. Experimental vs calculated melting point temperatures (°C) for 72 compounds, using linear multiple regression.

RESULTS AND DISCUSSION

Regression Model. The best model obtained by forward-MLR analysis to predict the MP temperatures of 72 1,2,3-diazaborines and the correlation coefficients between the molecular descriptors and the dependent variable are shown in Tables 2 and 3, respectively. The change in statistics during the successive addition of descriptors into the model is presented in Table 4. Figure 2 is a graphic of the predicted versus experimental MP temperature values. Descriptors in the best model were the following: sum of the atomic charges of the heteroatoms (ΣAC_{HET}), sum of the connectivity indexes $^0X^0$, $^0X^1$, and $^0X^2$ (Σ^0X), number of atoms in the molecule (NA), and the volume of the box (V_{BOX}). These descriptors have been shown as important in describing physicochemical parameters.[45] Results from the analysis of variance presented in Table 2 showed that the developed model was highly significant ($p < 0.0001$), suggesting that the chosen molecular descriptors were successful for predicting their MP temperatures. According to the significance levels calculated from the T-statistic values (Table 2) obtained for each descriptor, and the first parameter selected by the model (Table 4), the most important descriptor in the MLR model was (Σ^0X). The multicollinearity among descriptors was tested using both VIF and the individual correlation coefficient between them. Calculated VIF values were less than 10 (Table 2), and the correlation coefficient between pairs of descriptors, as shown in Table 5, did not have absolute values greater than 0.90, suggesting that there were not serious multicollinearity among descriptors.

Table 5. Correlation Matrix for the Descriptors Used in the Model

	ΣAC_{HET}	Σ^0X	NA	V_{BOX}
ΣAC_{HET}	1			
Σ^0X	0.32	1		
NA	−0.03	−0.86	1	
V_{BOX}	−0.34	−0.81	0.60	1

Table 6. Eigenvalues, Percentage of Variance Explained, and Cumulative Percentage of the Variance Reproduced by the Factor Analysis for the Descriptors Used in the Model

factor number	eigenvalue	% total variance	cumulative percentage
1	2.409	60.2	60.2
2	1.046	26.2	86.4
3	0.480	12.0	98.4
4	0.065	1.6	100.0

Table 7. Varimax Rotated Factor Matrix for the Descriptors Used To Predict the Melting Point Temperatures of 1,2,3-Diazaborines

descriptor	factor 1	factor 2
ΣAC_{HET}	0.071	−0.902
Σ^0X	0.811	0.547
NA	0.503	0.736
V_{BOX}	0.937	−0.077

Factor analysis was used to determine general multivariate relationships between selected variables in the predicted model.[50] The results are shown in Tables 6 and 7. Two factors had eigenvalues greater than one, and consequently data variability could be expressed in terms of two new reduced group of variables named factors. The first factor explained 60.2% of the total variance and the second one 26.2%, yielding a total of 86.4%. To have an easier data interpretation from calculated factors, the Varimax Factor Matrix was rotated, and the results were presented in Table 7. The analysis suggested that factor 1 was loaded by both V_{BOX} and Σ^0X, two descriptors associated with molecular size and shape. The second factor was linked to both ΣACHet and NA, descriptors that condense information on the electronic environment in the molecule. Therefore, both topological and electronic descriptors determine MP temperatures for 1,2,3-diazaborine molecules.

The four-descriptor model found for the MP temperatures of 1,2,3-diazaborines basically combines three types of molecular descriptors: topological, geometrical, and electronic. These results are in agreement with the findings reported by Hosoya et al.,[51] who considers very unlikely that the MP temperature could be defined from a simple QSAR relationship, using exclusively indices derived from the topological descriptors of individual molecules.

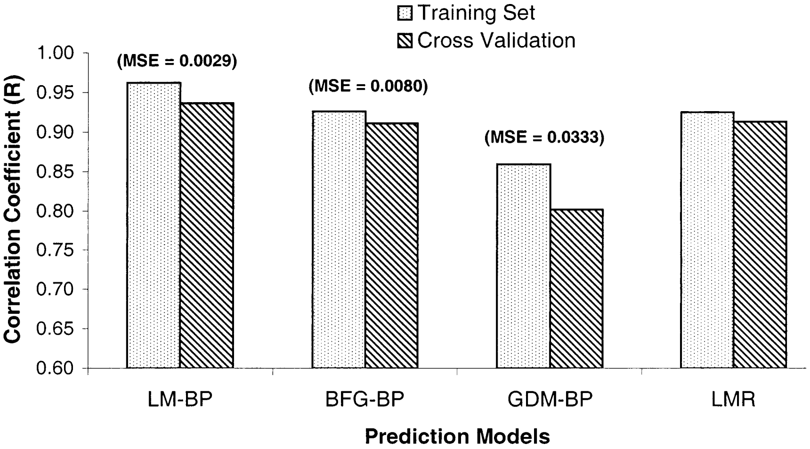


Figure 3. Comparison between correlation coefficients obtained for training (best neural netwok model) and cross-validation sets used during the application of three different BP algorithms with five hidden layers. Results for LMR are also presented. MSE values for different neural network models are given in parenthesis. LM = Levenberg−Marquardt, BFG = Quasi-Newton, GDM = gradient descent momentum.

Neural Network Model. The BP architecture was selected adjusting the parameters optimized by trial and error. Parameters such as learning rate and momentum coefficient were set before training to reduce the network error. The number of neurons in the input layer was four, one for each descriptor obtained in the MLR model, and the output layer contained one neuron, which corresponded to the vector of MP temperature values, as mentioned above. The number of nodes in the hidden layer was determined running the network from 4:1:1 (four input neurons, one hidden neuron and one output neuron) to 4:5:1 (four input neurons, five hidden neurons and one output neuron). It was observed that a 4:5:1 network that has 31 adjustable parameters minimized the error without sacrificing network quality. To avoid random effects in the network, the ratio of observations to total adjustable parameters should be kept above 2.0.[27,52] In this study, this ratio was above the recommendable minimum value, 65/31, or 2.1.

The best neural network was found testing the network architecture of 4:5:1 with different BP algorithms, including gradient descent with momentum, quasi-Newton, and Levenberg−Marquardt. Once the net architecture was established, the network was trained to optimize the weight and bias values. To control the network overfitting during the training procedure, the training set and the error were monitored using the validation sets. The error of the validation sets decreased during the initial phase of the training. However, when the network began to overfit the data, the error on the validation sets began to rise and then the training was stopped for a specified number of iterations, and the weights and biases at the minimum of the validation error were recorded. The performance of the BP neural network was evaluated using the MSE. In this study, the best model was selected by choosing the weight and bias values that yielded the lowest MSE for the training process.

The correlation coefficient between observed and calculated MP temperatures for the training and cross-validation sets as well as the MSE for the training set, using the neural network BP algorithms (mean of five sets), together with the corresponding results from MLR are presented in Figure 3. It is observed that the best results were computed using the neural network BP with the Levenberg−Marquardt algorithm. A graph showing the observed versus calculated MP temperatures for the best Levenberg−Marquardt BP neural network prediction model is presented in Figure 4.

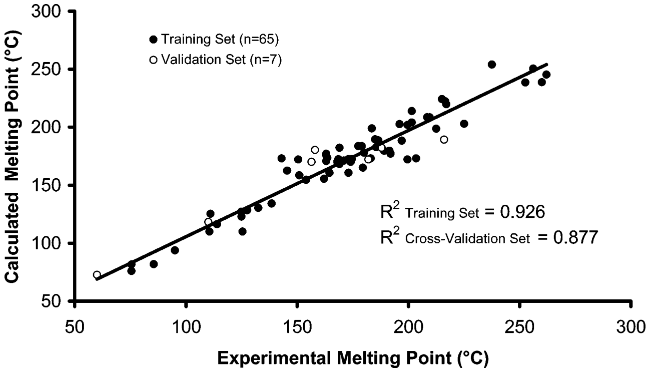


Figure 4. Experimental vs calculated melting points temperatures (°C) for a 65-compound training set and a seven-compound validation set using Levenberg−Marquardt BP neural network.

Descriptor Interpretation. MP temperature is a parameter that depends on the maximum capacity of the molecules to compact themselves within the crystal solid structure. This capacity is limited by molecular symmetry and intermolecular forces.[11] The summation of connectivity indexes (Σ^0X) is a topological descriptor related to the connectivity between atoms in the molecule and depends on molecular size. Although this topological descriptor describes the molecular graph, they have been extensively used as indicators of branching pattern.[53] It should be kept in mind that the conformations found on crystal structures can be affected by packing effects that are influenced by branching and the interactions generated with the rotations of the individual lateral groups. The second most important descriptor in the model was NA, variable that correlates with (Σ^0X) (R = 0.84). Although NA is a typical descriptor dependent on molecular size, it is also a function of molecular polarizability, an electronic parameter that encodes information about the dispersive interactions within the crystal system. The difference between these two descriptors is that Σ^0X encodes the molecular branching component.

It is remarkable that Σ^0X correlates with the rotational constant (RC) calculated from the optimized structure (R = −0.893, $P < 0.0001$). RC is defined as $h/(8\pi I)$, where h is the Planck's constant and I is the moment of inertia.[54] RC is used to calculate possible frequencies of pure-rotational spectral lines of molecules. The moment of inertia is dependent on atomic mass distribution along the main axis of the molecule. At the same time, the branching pattern determines the possibility of group rotation within the molecule and similarly, in this sense, like in a puzzle, the side chains can fit together in the crystal cell only when a particular conformation is achieved. A greater rotation probability will diminish packing capability and consequently the MP temperature will be lower.

The volume of the box (V_{BOX}) is a geometrical descriptor that implicitly encodes bulkiness and can be related to the crystal cell size. In the regression model this parameter has a negative regression coefficient. The inverse relationship between MP and V_{BOX} may indicate that an increase in crystal cell size decreases key electrostatic interactions between diazaborines and consequently the MP temperature is reduced.

Finally, the sum of the atomic charges for the heteroatoms ($\Sigma AC_{\mathbf{HET}}$) is an electronic descriptor determined basically by the atomic charges on the sulfonyl group (SO2). This charge distribution facilitates the formation of dipole−dipole interactions among diazaborines and must be a function of polarity as well as the molecular packing ability in the crystal solid. These charges on heteroatoms can also reflect hydrogen-bonding properties, which have been defined as important in the prediction of melting points for anilines.[55] At this point, it is important to emphasize that these data have been obtained considering molecules in the isolated state, and the intermolecular forces participating in the packing within the crystal solid were not considered.

CONCLUSION

The results of this study demonstrate that multiple-linear-regression analysis and BP neural network are techniques that can be used to predict successfully the MP temperatures of the 1,2,3-diazaborine compounds. The most accurate results were obtained using a neural network back-propagation optimized by the Levenberg−Marquardt algorithm. The current findings have allowed us to conclude that MP temperatures for these compounds appear to be governed by branching, electronic, and geometrical factors not absolutely independent one from another. Thus, the molecular crystal formation for 1,2,3-diazaborines includes a balance between intermolecular interactions and group packing processes within the cell crystal.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the National Institute of Science and Technology of Colombia (Colciencias), Bogotá, and the University of Cartagena, Cartagena (Colombia).

REFERENCES AND NOTES

(1) Davis, M. C.; Franzblau, S. G.; Martin, A. R. Syntheses and evaluation of benzodiazaborine compounds against M. tuberculosis H37Rv in vitro. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 843−846.

(2) Baldock, C.; de Boer, G. J.; Rafferty, J. B.; Stuitje, A. R.; Rice, D. W. Mechanism of action of diazaborines. *Biochem. Pharmacol.* **1998**, *55*, 1541−1549.

(3) Baldock, C.; Rafferty, J. B.; Sedelnikova, S. E.; Baker, P. J.; Stuitje, A. R.; Slabas, A. R.; Hawkes, T. R.; Rice, D. W. A mechanism of drug action revealed by structural studies of enoyl reductase. *Science* **1996**, *274,* 2107−2110.

(4) Heath, R. J.; White, S. W.; Rock, C. O. Inhibitors of fatty acid synthesis as antimicrobial hemotherapeutics. *Appl. Microbiol. Biotechnol.* **2002**, *58*, 695−703.

(5) Roujeinikova, A.; Sedelnikova, S.; de Boer, G. J.; Stuitje, A. R.; Slabas, A. R.; Rafferty, J. B.; Rice, D. W. Inhibitor binding studies on enoyl reductase reveal conformational changes related to substrate recognition. *J. Biol. Chem.* **1999**, *274*, 30811−30817.

(6) Jungwirth, H.; Bergler, H.; Hogenauer, G. Diazaborine treatment of Baker's yeast results in stabilization of aberrant mRNAs. *J. Biol. Chem.* **2001**, *276*, 36419−36424.

(7) Grassberger, M. A.; Turnowsky, F.; Hildebrandt, J. Preparation and antibacterial activities of new 1,2,3-diazaborine derivatives and analogues. *J. Med. Chem.* **1984**, *27*, 947−953.

(8) Lee, G. T.; Prasad, K.; Repič, O. A facile synthesis of 2,4-diaza-1-borines from anilines. *Tetrahedron Lett.* **2002**, *43*, 3255−3257.

(9) Oprea, T. I. On the information contend of 2D and 3D descriptors for QSAR. *J. Bruz. Chem. Soc***. 2002**, *13*(6), 811−815.

(10) Browm, R. D.; Martín, Y. C. Use of Structure−Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 572−584.

(11) Browm, R. D.; Martín, Y. C. The Information Content of 2D and 3D Structural Descriptors Relevant to Ligand−Receptor Binding. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1−9.

(12) Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). Effect of Shape in Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc*. **1988**, *110*, 5959−5967.

(13) Goodford, P. J. Computational procedure for determining Energetically Favorable Binding Sites on Biological Important Macromolecules. *J. Med. Chem*. **1985**, *28*, 849−857.

(14) Wold, S.; Johanson, E.; Cocchi, M. In *3D QSAR in drug design: Theory, Method and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 523−550.

(15) Argese, E.; Bettiol, C.; Giurin, G.; Miana, P. Quantitative structure−activity relationships for the toxicity of chlorophenols to mammalian submitochondrial particles. *Chemosphere* **1999**, *38*, 2281−2292.

(16) Benoit-Guyod, J. L.; Andre, C.; Taillandier, G.; Rochat, J.; Boucherle, A. Toxicity and QSAR of chlorophenols on Lebistes reticulatus. *Ecotoxicol. Environ. Saf.* **1984**, *8*, 227−235.

(17) Dearden, J. C. The QSAR prediction of melting point, a property of environmental relevance. *Sci. Total Environ.* **1991**, *109−110*, 59−68.

(18) Katritzky, A. R.; Jain, R.; Lomaka, A.; Petrukhin, R.; Maran, U.; Karelson, M. Perspective on the relationship between melting points and chemical structure. *Crystal Growth Design* **2001**, *1*, 261−265.

(19) Zhao, L.; Yalkowsky, S. H. A combined group contribution and molecular geometry approach for predicting melting points of aliphatic compounds. *Ind. Eng. Chem. Res.* **1999**, *38*, 3581−3584.

(20) Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. Prediction of melting points for the substituted benzenes: A QSPR approach. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 913−919.

(21) Katritzky, A. R.; Lomaka, A.; Petrukhin, R.; Jain, R.; Karelson, M.; Visser, A. E. Rogers RD. QSPR correlation of the melting point for pyridinium bromides, potential ionic liquids. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 71−74.

(22) Levy, C. W.; Baldock, C.; Wallace, A. J.; Sedelnikova, S.; Viner, R. C.; Clough, J. M.; Stuitje, A. R.; Slabas, A. R.; Rice, D. W.; Rafferty, J. B. A study of the structure−activity relationship for diazaborine inhibition of *Escherichia coli* enoyl-ACP reductase. *J. Mol. Biol.* **2001**, *309*, 171−180.

(23) Ball, J.; Jurs, P. C. Automated selection of regression models using neural network for 13C NMR spectral predictions. *Anal. Chem.* **1993**, *65*, 505−512.

(24) Li, Z.; Cheng, Z.; Xu, L.; Li, T. Nolinear fitting using a neural net algorithm. *Anal. Chem.* **1993**, *65*, 393−396.

(25) Xu, L.; Ball, J. W.; Dixon, S. L.; Jurs, P. C. Quantitative structure−activity relationships for toxicity of phenols using regression analysis and computational neural network. *Environ. Toxicol. Chem*. **1994**, *13*(5), 841−851.

(26) Wikel, J. Dow, E. R.; Heathman, M. Interpretative Neural Networks for QSAR. http://www.netsci.org/Science/Combichem/feature02.html (accessed in March 2003).

(27) Wessel, W. D.; Sutter, J. M.; Jurs, P. C. Prediction of reduced ion mobility constants of organic compounds from molecular structure. *Anal. Chem*. **1996**, *68*(23), 4237−4243.

(28) Breindl, A.; Beck, B.; Clarck, T. Prediction of the *n*-Octanol/Water Partition Coefficient, logP, Using a Combination of Semiempirical MO−Calculations and a Netral Network. *J. Mol. Model.* **1997**, *3*, 142−145.

(29) Nestorov, I.; Rowland, M. Empirical versus mechanistic modeling: Comparison of an Artificial Neural Network to a mechanistically based model for quantitative structure pharmacokinetic relationships of a homologous series of barbiturates. *AAPS Pharmsci.* **1999**, *1*(4), 1−8.

(30) Jalali-Heravi, M.; Fatemi, M. H. Artificial neural network modeling of kovats retention indices for nocyclic and monocyclic terpenes *J. Chromatogr. A.* **2001**, *915*, 177−183.

(31) Jalali-Heravi, M.; Garkani-Nejad, Z. Use of self- training artificial neural networks in modeling of gas chromatographic relative retention times of a variety of organic compounds. *J. Chromatogr. A.* **2002**, *945*, 173−184

(32) Fatemi, M. H. Simultaneous modeling of Kovats retention indices on OV-1 and SE−54 stationary phases using artificial neural network. *J. Chromatogr. A* **2002**, *955*(2), 273−280.

(33) Baskin, I. I.; Ait, A. O.; Halberstam, N. M.; Palyulin, V. A.; Zefirov, N. S. An approach to the interpretation of back-propagation neural network models in QSAR studies. SAR QSAR *Environ. Res.* **2002**, *13*(1), 35−41.

(34) Mosier, P. D.; Jurs, P. C. QSAR/QSPR studies using probabilistic neural networks and generalized regression neural networks. *J. Chem. Inf. Comput. Sci.* **2002**, *42*(6), 1460−1470.

(35) Niwa, T. Using general regression and probabilistic neural networks to predict human intestinal absorption with topological descriptors derived from two-dimensional chemical structures. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 113−119.

(36) Demuth, H.; Beale, M. *Neural Network Toolbox For Use with MATLAB User's Guide Version 3.0*; 5th ed.; The MathWorks, Inc.: 1998.

(37) Jansson. P. A. Neural network: An Overview. *Anal. Chem*. **1991**, *63*(6), 357A−362A.

(38) Katritzky, A. R.; Murugan, R.; Grendze, M. P.; Toomey, J. E.; Karelson, M., Jr.; Lobanov, V.; Rachwal. P. Predicting Physical Properties from Molecular Structure. *Chemtech*. **1994**, *24*, 17−23.

(39) Stanton, D. T. Development of a quantitative structure−property Relationship model for estimating Normal Boiling Point of Small Multifunctional Organic Molecules. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 81−90.

(40) SPARTAN. Wave Function, Inc. 1997.

(41) Gaussian 94, Revision D.3; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 1995.

(42) Olivero, J.; Payares, P.; Diaz, D.; Vivas, R.; Mercado, J. PCDM V.2. 1995. Modified by Pacheco, L.; Johnson, B. University of Cartagena, Cartagena, Colombia, South America, 2000.

(43) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.

(44) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Chemistry and Drug Research*; Bawden, D., Ed.; Research Studies Press LTD.: Letchworth, Hertfordshire, England, 1986; pp 1−24.

(45) Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835−857.

(46) Statgraphics Plus for Windows. Statistical graphics System, User's Guide version 3.0. Statistical graphics corporation. 1999.

(47) Alvarez, R. *Estadística Multivariante y no paramétrica con SPSS: aplicación a las ciencias de la salud*; Diaz de Santos, Eds.; Madrid, Spain, 1995.

(48) Woloszyn, T. F.; Jurs, P. C. Quantitative structure-retention relationship studies of sulfur vesicants. *Anal. Chem*. **1992**, *64*, 3059−3063.

(49) Bakken, G. A.; Jurs, P. C. Prediction of Metihyl Radical Addition Rate Constants from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 508−514.

(50) Afifi, A.; Clark, V. *Computer-aided multivariate analysis*; Reinhold, V. N., Ed.; New York, 1990.

(51) Hosoya, H.; Gotoh, M.; Murakami, M.; Ikeda, S. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkenes with the topology of graphs? *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 192−196.

(52) MacElroy, N. R.; Jurs, P. C. Prediction of Aqueous Solubility of Heteroatom- Containing Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1237−1247.

(53) Pompe, M.; Novic, M. Prediction of gas-chromatographic retention indices using topological descriptors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 59−67.

(54) Levine, I. N. *Quantum Chemistry*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, 2000.

(55) Dearden, J. C.; Ghafourian, T. Hydrogen bonding parameters for QSAR: comparison of indicator variables, hydrogen bond counts, molecular orbital and other parameters. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 231−235.

CI020387K