Structure-Based Predictions of ^{1}H NMR Chemical Shifts Using Feed-Forward Neural Networks

Yuri Binev† and João Aires-de-Sousa*

REQUIMTE, CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal

Received October 22, 2003

Feed-forward neural networks were trained for the general prediction of ^{1}H NMR chemical shifts of CH_n protons in organic compounds in $CDCl_3$. The training set consisted of 744 ^{1}H NMR chemical shifts from 120 molecular structures. The method was optimized in terms of selected proton descriptors (selection of variables), the number of hidden neurons, and integration of different networks in ensembles. Predictions were obtained for an independent test set of 952 cases with a mean average error of 0.29 ppm (0.20 ppm for 90% of the cases). The results were significantly better than those obtained with counterpropagation neural networks.

INTRODUCTION

Accurate estimation of NMR chemical shifts from molecular structure plays an important role in structure elucidation and structure confirmation. Furthermore, predictions must be produced fast and automatically to be useful for automatic structure elucidation[1] or for the analysis of large sets of samples.[2] In this context, ab initio calculations are too time-consuming, and for most organic compounds the accuracy obtained is only comparable to faster empirical methods based on large amounts of experimental data.

Although several successful approaches have been described for ^{13}C NMR chemical shifts,[3] the area of ^{1}H NMR has been more problematic and less explored, for reasons probably related to experimental data (less available and more difficult to process) and the physics of NMR itself (more complex relationships between structure and properties and higher dependence on the experimental conditions).[4]

Database search and increment rules have been the first methods to be implemented. In database-centered methods, a query proton belonging to a specific structure is submitted, and a database of atom structural environments (usually represented by HOSE codes[5]) is searched. The prediction of the chemical shift is then based on the chemical shifts of the most similar atoms found. Such a method is used in commercial packages such as ACD[6] or Chemical Concepts SpecInfo.[7]

With increment rules,[8,9] predictions are based on tabulated chemical shifts for classes of structures and corrected with additive contributions from neighboring functional groups or substructures. Several tables have been compiled for different types of nuclei.[10−12] Examples of such implementations include ChemDraw (CambridgeSoft)[13] and TopNMR (Upstream).[14] It has been claimed that the incremental methods can predict the ^{1}H NMR shifts of about 90% of all CH_n groups with a mean deviation between 0.2 and 0.3 ppm.[14]

More recently, neural networks were trained with a small data set of protons from 120 structures and have been used with success for the fast estimation of NMR chemical shifts of CH_n protons.[15] The relationship between protons in defined molecular structures and the corresponding ^{1}H NMR chemical shift was established by counterpropagation neural networks (CPG NN), which used descriptors for hydrogen atoms as input and the chemical shift of the corresponding proton as output. Various types of descriptors were used, namely, topological, physicochemical, and geometric descriptors. Four different classes of protons were treated separately: protons belonging to aromatic systems, protons belonging to π nonaromatic systems, protons belonging to rigid aliphatic substructures, and protons belonging to nonrigid aliphatic substructures. Some descriptors were common to all four classes; others were specific for a single class. When applied to the prediction of 259 chemical shifts from 31 molecules (an independent test set), the mean absolute error obtained for the whole prediction set was 0.25 ppm, and for 90% of the cases the mean absolute error was 0.19 ppm.

Employing exactly the same data for training, we could now significantly improve the predictions by using ensembles of *feed-forward* neural networks (FFNNs) and optimizing several factors, including the selection of variables and the architecture of the networks. Ensembles of neural networks were employed, because different nets, trained in order to be consistent with the training data, can give different predictions for new examples. Ensembles of neural networks offer a possible solution to this problem by making predictions on the basis of "votes" from individual networks.[16] They yield more stable (and often more accurate) predictors than individual networks.[17] In this article, the optimization of the feed-forward neural networks is described, and the predictions obtained for a much larger test set of 952 cases are reported.

* Corresponding author phone: +351-21-2948575; fax: +351-21-2948550; e-mail: jas@fct.unl.pt.

† Permanent address: Institute of Organic Chemistry, Bulgarian Academy of Sciences, BG-1113 Sofia, Bulgaria.

10.1021/ci034228s CCC: $27.50
Published on Web 02/13/2004

METHODOLOGY

Data Sets and Classification of Hydrogen Atoms. The training and prediction sets used in the previous work[15] with CPG neural networks were now used again. The training set consists of 744 ^{1}H NMR chemical shifts for protons from 120 molecular structures. The prediction set (here named "prediction set A") contains 259 chemical shifts from 31 molecules. In this work, a larger prediction set (prediction set B) was also used to test the methods. This is a heterogeneous set, obtained from 100 structures, and includes 952 experimental chemical shifts with the corresponding hydrogen atoms. Because stereochemistry is not defined in prediction set B, when two hydrogen atoms are bonded to the same carbon atom, their chemical shifts are averaged.

Only data from spectra obtained in $CDCl_3$ were considered. The collection was restricted to CH_n protons and to compounds containing elements C, H, N, O, S, F, Cl, Br, or I. Chemical shifts of protons bonded to heteroatoms were not included since they are strongly influenced by experimental conditions (such as the concentration of the sample).

Four different classes of protons were defined (aromatic, nonaromatic π, rigid aliphatic, and nonrigid aliphatic), and protons belonging to each of them were treated separately. This procedure allowed the use of more specific descriptors for each class. Protons were classified as (a) aromatic when they are bonded to an aromatic system, (b) nonaromatic π when they are bonded to a nonaromatic π system, (c) rigid aliphatic when a nonrotatable bond is identified in the second sphere of bonds centered on the proton, and (d) nonrigid aliphatic when not included in previous classes. A bond was defined as nonrotatable if it belongs to a ring, to a π system, or to an amid functional group. The training set was divided accordingly into four training sets (one for each class), and the same procedure was applied to the prediction sets. The number of cases in the training set and prediction set A were respectively 145 and 39 for aromatic protons, 117 and 47 for π protons, 237 and 87 for rigid aliphatic protons, and 245 and 86 for nonrigid aliphatic protons. The number of protons in prediction set B were respectively 247 for aromatic protons, 93 for π protons, 237 for rigid aliphatic protons, and 375 for nonrigid aliphatic protons. The experimental chemical shifts were linearly normalized between 0.1 and 0.9 within each class, on the basis of the training set.

Each training set was further divided into a cross-validation set and a reduced training set with approximately the same sizes, the division being done exactly as in the work with CPG NNs.[15] These are the *initial cross-validation sets* and the *initial reduced training sets*.

Descriptors of Hydrogen Atoms. To be submitted to a neural network, each proton was represented by a fixed number of descriptors. For hydrogen atoms belonging to aromatic, rigid aliphatic, and nonrigid aliphatic classes, the same descriptors were used as in ref 15 and are described there. These include physicochemical, geometric, and topological descriptors. Physicochemical descriptors were based on empirical values calculated by the software package PETRA[18] (version 3.0) comprising a variety of published methods[19,20] for the protons and for their neighborhood. Examples of physicochemical descriptors used in this study are as follows: partial atomic charge of the proton, effective polarizability of the proton, average of partial atomic charges of atoms in the second sphere, maximum partial atomic charges of atoms in the second sphere, minimum effective polarizability of atoms in the second sphere, and the average of σ electronegativities of atoms in the second sphere.

Geometric descriptors were based on the 3D molecular structure generated by the CORINA software package.[21−23] The global 3D environment of a proton belonging to a rigid substructure was represented by a radial distribution function (RDF).[24]

Topological descriptors were based on the analysis of the connection table. Some examples of topological descriptors are as follows: the number of carbon atoms in the second sphere centered on the proton, the number of oxygen atoms in the third sphere, and the number of atoms in the second sphere that belong to an aromatic system. A topological radial distribution function was also used, where the sum of bond lengths on the shortest possible path between the proton and another atom is used instead of 3D interatomic distances.[15] All in all, 92 descriptors were used for aromatic protons, 119 for nonrigid aliphatic protons, and 174 for rigid aliphatic protons.

For nonaromatic π protons, some new descriptors were now introduced to replace some of the old. The new descriptors discriminate properties of atoms at *gem*, cis, and trans positions relative to the proton. Some of the old descriptors were the average, maximum, and minimum of properties of atoms at a distance of two bonds from the proton, which included the *gem* atom. Other descriptors were the average, maximum, and minimum of properties of atoms at a distance of three bonds, which included the cis and trans atoms together. Now, for example, one descriptor is the charge of the atom at the cis position and another descriptor is the charge of the atom at the trans position, instead of the average of the charges of atoms three bonds away from the proton. An expanded set of 110 descriptors resulted.

The full list of descriptors is available in the Supporting Information.

Every descriptor, except those obtained from RDF functions, was linearly normalized between 0.1 and 0.9, within each class, on the basis of the training sets.

Feed-Forward Neural Networks. Fully connected feed-forward neural networks were trained with the back-propagation of errors algorithm. They were implemented with in-house developed software based on JATOON.[25,26] The network architecture consisted of an input layer (including a bias equal to 1), one hidden layer (also including a bias equal to 1), and one output neuron. The networks were trained with the examples from the reduced training sets. They were trained to produce the (normalized) chemical shift value as the output when the descriptors representing a proton were submitted as input. Before starting the training of each network, the training set was randomly divided into equal-size cross-validation set and reduced training set. This procedure was particularly important when ensembles of networks were trained with the same data (see below). The training was performed until (a) a minimum root-mean-square of errors (RMSE) was found for the cross-validation set and (b) a new minimum could not be found after a number of epochs (typically 500 epochs). After training, the networks were tested with independent sets.

Table 1. Mean Absolute Errors (ppm) of the Predictions of ^{1}H NMR Chemical Shifts Obtained by Counterpropagation Neural Networks (CPG NN) and Feed-Forward Neural Networks (FFNN) Using All of the Descriptors

	prediction set A		prediction set B	
class	CPG NN	FFNN	CPG NN	FFNN
aromatic	0.23	0.23	0.29	0.26
nonaromatic π	0.22	0.20	0.43	0.38
nonrigid aliphatic	0.21	0.14	0.24	0.18
rigid aliphatic	0.34	0.32	0.51	0.44

For the optimization of the size of NN ensembles (see below) the ASNN program[27] was used to train the networks, with the Levenberg−Marquardt algorithm.[28]

Ensembles of FFNNs. An ensemble of FFNNs is a set of networks, independently trained, which contribute to a single prediction.[16,17] The prediction is obtained as the average of the outputs from the FFNNs of the ensemble. As mentioned above, prior to the training of each network, the training set was randomly divided into new cross-validation and reduced training sets with approximately the same sizes. In this way, each network of the ensemble was trained with its own training and cross-validation sets, in addition to being seeded with its own (random) initial weights.

Selection of Descriptors. Descriptors were selected by stepwise removal of correlations. The descriptors were sorted in decreasing order of the Pearson correlation coefficients when used in a global search for two-descriptor correlations and removed stepwise according to the obtained order. Experiments were then performed, using a decreasing number of descriptors to train the neural networks. In these experiments, the number of hidden neurons was set to one-third of the number of descriptors. Twelve networks were trained (ensemble of 12 NNs) with each subset of descriptors, and predictions were obtained as the average of the 12 predictions. The smallest possible number of descriptors giving rise to the best predictions for the initial cross-validation set was chosen as the optimum.

Optimization of the Number of Hidden Neurons. Using the selected descriptors, FFNNs were trained with different numbers of hidden neurons. In each experiment, 12 networks with the same number of hidden neurons were trained, and the predictions were obtained as the average of the 12 predictions. Different experiments were performed with the number of hidden neurons varying between 3 and one-third the number of descriptors. The choice of the optimum number of hidden neurons was based on the errors of the predictions for the initial cross-validation set.

Optimization of the Size of FFNN Ensembles. Ensembles with 12, 25, 50, and 100 FFNNs were tried. The decision about the best size of the ensembles took into account the predictions obtained for prediction set A, the convenience of large ensembles for posterior use as associative neural networks[26] (see the following paper in this issue), but at the same time the inconvenience of too large ensembles in terms of computational requirements.

RESULTS AND DISCUSSION

In a first experiment with FFNNs using all the available descriptors, improved predictions could be obtained in comparison to the counterpropagation NNs (Table 1). This was particularly true for the larger test set (prediction set B). The architecture of FFNNs makes them generally better at adjusting quantitative relationships between the input and the output values. It was also found that the new descriptors for the nonaromatic π protons slightly decreased the mean average errors (old descriptors, 0.20 ppm; new descriptors, 0.17 ppm, prediction set A). These new descriptors for nonaromatic π protons included properties of atoms specifically at *gem*, trans, and cis positions relative to the protons (see Methodology).

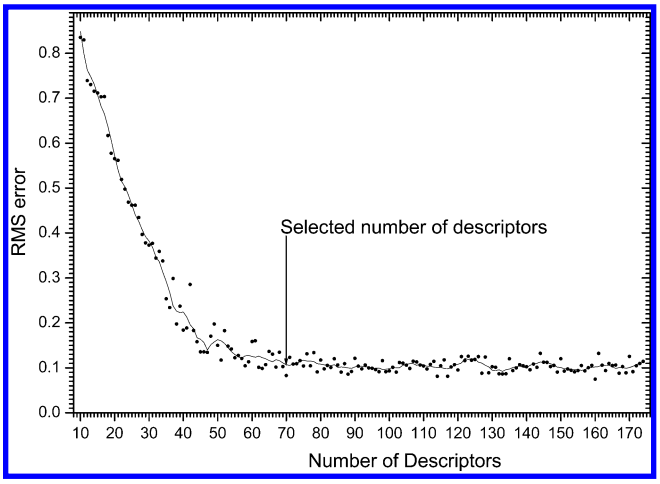

Figure 1. Influence of the number of selected descriptors on the errors obtained for the cross-validation set, in the case of rigid protons.

Table 2. Mean Absolute Errors (ppm) of the Predictions of ^{1}H NMR Chemical Shifts for Prediction Set A Obtained by FFNNs Using All Descriptors and Selected Descriptors

class	all descriptors (no. of descriptors)	selected descriptors (no. of descriptors)
aromatic	0.23 (92)	0.21 (50)
nonaromatic π	0.17 (110)[a]	0.22 (61)
nonrigid aliphatic	0.14 (119)	0.14 (53)
rigid aliphatic	0.32 (174)	0.35 (70)

[a] Using the new descriptors for nonaromatic π protons. See Methodology.

The selection of descriptors was performed by stepwise removal of interdescriptor correlations, and it was done for each of the four classes of protons independently. The choice of the best number of descriptors took into account the quality of the predictions for the cross-validation set, and also the convenience of compact models. An example of the influence of the number of descriptors on the errors is shown in Figure 1, for the rigid protons. With this procedure, it was possible to remove more than half of the descriptors, maintaining the excellent performance of the networks (Table 2). This indicates that the original set of descriptors contained a high degree of redundancy. The selected descriptors are qualitatively described for each case in Table 3. The full list is available in the Supporting Information.

With the descriptors selected as described above, the influence of the number of hidden neurons on the quality of predictions was investigated. The best found numbers of hidden neurons for aromatic, nonaromatic π, rigid aliphatic, and nonrigid aliphatic classes were respectively 9, 6, 10, and 10.

Table 4 shows the results obtained with ensembles of FFNNs of different sizes. It also shows the predictions from

Table 3. Selected Descriptors for the Four Classes of Protons

class	topological descriptors	physicochemical descriptors	geometrical descriptors	topological RDF codes	total	initial set
aromatic	10	21	1	18	50	92
nonaromatic π	13	21	9	14	61	110
nonrigid aliphatic	10	22	0	21	53	119
rigid aliphatic	8	8	43	11	70	174

Table 4. Mean Absolute Errors (ppm) of the Predictions of ^{1}H NMR Chemical Shifts Obtained for Prediction Set A by Ensembles of FFNNs with Different Sizes or by Individual FFNNs

	12 FFNNs	25 FFNNs	50 FFNNs	100 FFNNs	best individual FFNN	worst individual FFNN
aromatic	0.21	0.20	0.21	0.20	0.17	0.30
nonaromatic π	0.25	0.23	0.22	0.22	0.16	0.51
nonrigid aliphatic	0.18	0.15	0.15	0.16	0.15	0.24
rigid aliphatic	0.35	0.34	0.34	0.34	0.35	0.47

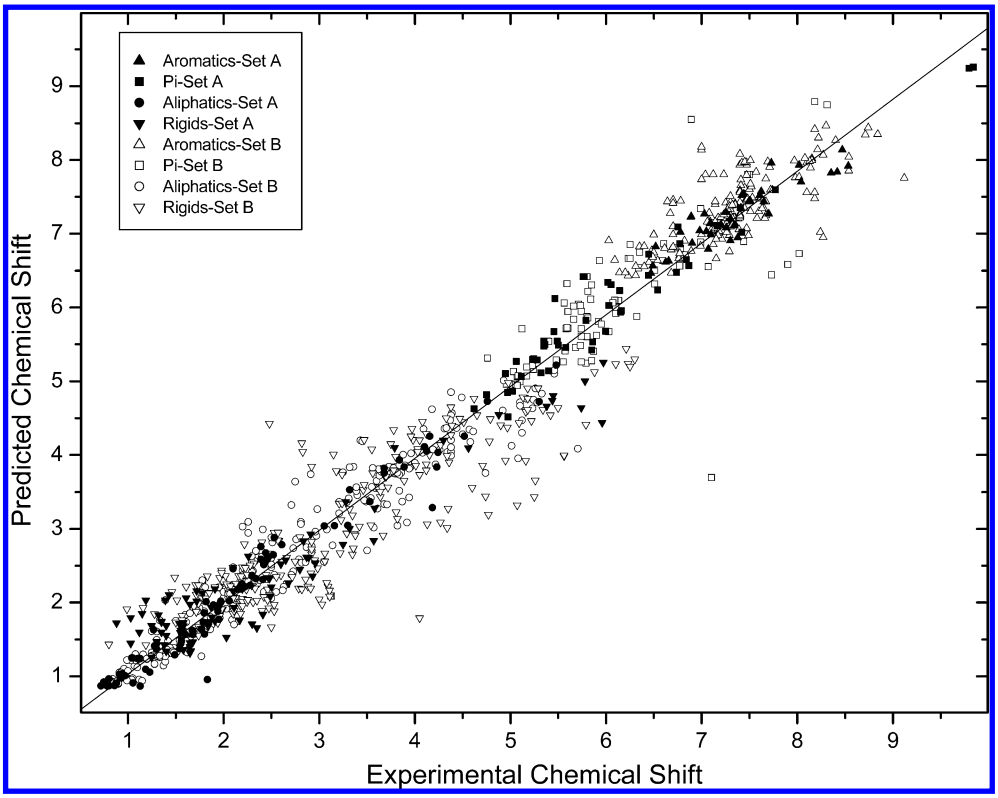

Figure 2. Experimental ^{1}H NMR chemical shifts vs predictions from optimized FFNNs for prediction sets A and B ($R^2 = 0.9704$).

the single best network and the single worst network of the ensembles. The errors decreased with an increasing number of networks up to 25 networks and generally stabilized after that. It was then decided to use ensembles of 50 FFNNs for all of the classes, except the class of rigid aliphatic protons, which require more descriptors than the others, and a smaller ensemble of 25 FFNNs was chosen. In all four classes, it can be seen that the ensembles always work much better than the worst single network. In several situations, the error of the chosen ensemble is very close to the error of the best single network. For the rigid aliphatic protons, the chosen ensemble of 25 FFNNs works even better than the best single network. The use of an ensemble of networks hence had the advantage of more stable results, and higher robustness is expected in predictions for new data sets. In fact, it was confirmed that if the best individual network for prediction set A is applied to prediction set B, the error comes higher than with the ensemble in every type of protons.

Table 5. Mean Absolute Errors (ppm) of the Predictions of ^{1}H NMR Chemical Shifts Obtained by Optimized FFNNs and CPG NNs

	prediction set A			prediction set B		
class	CPG NN	FFNN	no. of cases	CPG NN	FFNN	no. of cases
aromatic	0.21	0.21	39	0.29	0.26	247
nonaromatic π	0.22	0.22	47	0.46	0.35	93
nonrigid aliphatic	0.19	0.15	86	0.24	0.19	375
rigid aliphatic	0.34	0.34	87	0.55	0.43	237
total	**0.25**	**0.24**	**259**	**0.36**	**0.29**	**952**

After all the optimizations, the final results for prediction sets A and B are summarized in Table 5, together with the best results from counterpropagation NNs (the descriptors used in CPG NNs had been selected by a more appropriate method for that type of network—genetic algorithms—which is described in ref 15). The individual predictions from the FFNNs are plotted in Figure 2 against the experimental

chemical shifts. Globally the predictions of FFNNs are much more accurate than those of CPG NNs, particularly for prediction set B. However, the largest individual error in prediction set B was larger with FFNNs than with CPG NNs. This was the case of a hydrogen atom bonded to a cyclopropene, which is a structure not covered by the training data. When applied outside of their training domain (extrapolation), FFNNs can give disparate predictions, while CPG NNs never really extrapolate as they basically work as look-up tables. This can explain the observation. The use of associative neural networks (see the following paper in this issue) offers a solution to this problem.

Predictions for the class of rigid aliphatic protons were generally more problematic. It can be explained by the highly complex influence of 3D factors on the chemical shift. In this class, the diversity of possible substructures encompasses not only all the combinations of neighbor functional groups but also a highly diverse range of rigid 3D environments. For this wide space, the small available training set is clearly very narrow.

In prediction set B, stereochemistry of two nonequivalent CH_2 protons was not assigned, although the database sometimes lists two different chemical shifts for two protons bonded to the same carbon atom. This fact was relevant for nonaromatic π and for rigid aliphatic protons, because diastereotopic protons are represented by nonidentical descriptors, and therefore the networks predict generally different chemical shifts. To overcome this limitation of data, the predictions for the two diastereotopic protons of CH_2 groups were averaged and compared to the similarly computed average of experimental values.

This study only used spectra taken in $CDCl_3$. However, it does not mean the method cannot be applied to NMR in other solvents. If networks are trained and optimized with experimental data obtained in another solvent, in principle they should be able to make predictions for chemical shifts in that solvent.

CONCLUSIONS

The application of FFNNs significantly improved the predictions of ^{1}H NMR chemical shifts by comparison to CPG NNs. The improvement was more dramatic for prediction set B. Removal of highly correlated proton descriptors and optimization of the number of hidden neurons produced much more compact models, maintaining a high level of accuracy. Ensembles of neural networks showed higher robustness and are expected to give better predictions in new situations. With the optimized system of networks, a global average mean error for an independent test set of 952 cases was 0.29 ppm and for 90% of the cases 0.20 ppm.

Remarkable results could be achieved with relatively small training sets, possibly due to the use of physicochemical and topological descriptors that generalize atom types to inherent physicochemical properties.

The optimized models were implemented in a web service for fast estimation of ^{1}H NMR chemical shifts, which is freely available at http://www.dq.fct.unl.pt/spinus and http://www2.chemie.uni-erlangen.de/services/spinus.

ACKNOWLEDGMENT

The authors are greatly indebted to Prof. Johann Gasteiger for his continuous support, and access to software developed by his research group. The authors also thank Dr. Igor Tetko who made available the ASNN program. Y.B. acknowledges Fundação para a Ciência e Tecnologia (Lisbon, Portugal) for a postdoctoral grant under the POCTI program (SFRH/BPD/7162/2001). Chemical Concepts GmbH is acknowledged for sharing the data included in prediction set B.

Supporting Information Available: Full list of proton descriptors and list of selected descriptors for each class of protons. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Steinbeck, C. Computer-Assisted Structure Elucidation. In *Handbook of Chemoinformatics*; Gasteiger, J., Engel, T., Eds.; Wiley-VCH: New York, 2003; Vol. 3, Chapter 2.3, pp 1378−1406.

(2) For an example of combinatorial library analysis by ^{1}H NMR, see: Kalelkar, S.; Dow, E. R.; Grimes, J.; Clapham, M.; Hu, H. Automated Analysis of Proton NMR Spectra from Combinatorial Rapid Parallel Synthesis Using Self-Organizing Maps. *J. Comb. Chem.* **2002**, *4* (6), 622−629.

(3) For a recent example, and useful references, see: Meiler, J.; Meusinger, R.; Will, M. Fast Determination of C-13 NMR Chemical Shifts Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1169−1176.

(4) Steinbeck, C. Correlations between Chemical Structures and NMR Data. In *Handbook of Chemoinformatics*; Gasteiger, J., Engel, T., Eds.; Wiley-VCH: New York, 2003; Vol. 3, Chapter 2.2, pp 1368−1377.

(5) Bremser W. HOSE−A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355−365.

(6) Advanced Chemistry Development, Inc., http://www.acdlabs.com.

(7) Chemical Concepts GmbH, http://www.chemicalconcepts.com.

(8) Schaller, R. B.; Pretsch, E. A Computer Program for the Automatic Estimation of ^{1}H NMR Chemical Shifts. *Anal. Chim. Acta* **1994**, *290*, 295−302.

(9) Fürst, A.; Pretsch, E. A Computer Program for the Prediction of ^{13}C NMR Chemical Shifts of Organic Compounds. *Anal. Chim. Acta* **1990**, *229*, 17−25.

(10) Fürst, A.; Robien, W.; Pretsch, E. A Comprehensive Parameter Set for the Prediction of the ^{13}C NMR Chemical Shifts of sp^3-Hybridized Carbon Atoms in Organic Compounds. *Anal. Chim. Acta* **1990**, *233*, 213−222.

(11) Pretsch, E.; Fürst, A.; Robien, W. Parameter Set for the Prediction of the ^{13}C NMR Chemical Shifts of sp^2- and sp-Hybridized Carbon Atoms in Organic Compounds. *Anal. Chim. Acta* **1991**, *248*, 415−428.

(12) Schaller, R. B.; Arnold, C.; Pretsch, E. New Parameters for Predicting ^{1}H NMR Chemical Shifts of Protons Attached to Carbon Atoms. *Anal. Chim. Acta* **1995**, *312*, 95−105.

(13) CambridgeSoft Corp., http://www.cambridgesoft.com.

(14) Upstream Solutions GmbH, http://www.upstream.ch.

(15) Aires-de-Sousa, J.; Hemmer, M.; Gasteiger, J. Prediction of H-1 NMR Chemical Shifts Using Neural Networks. *Anal. Chem.* **2002**, *74*, 80−90.

(16) Dietterich, T. G. Ensemble Learning. In *The Handbook of Brain Theory and Neural Networks*; Arbib, M. A., Ed.; MIT Press: Cambridge, MA, 2002; pp 405−408.

(17) Agrafiotis, D. K.; Cedeno, W.; Lobanov, V. S. On the Use of Neural Network Ensembles in QSAR and QSPR. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 903−911.

(18) PETRA can be tested on the website http://www2.chemie.uni-erlangen.de and is available from Molecular Networks GmbH, http://www.mol-net.de.

(19) Gasteiger, J. Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds. In *Physical Property Prediction in Organic Chemistry*; Jochum, C., Hicks, M. G., Sunkel, J., Eds.; Springer-Verlag: Heidelberg, Germany, 1988; pp 119−138.

(20) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity−A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(21) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.

(22) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1992**, *3*, 537−547.

(23) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.

(24) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. The Prediction of the 3D Structure of Organic Molecules from Their Infrared Spectra. *Vibrat. Spectrosc.* **1999**, *19*, 151−164.
(25) Aires-de-Sousa, J. JATOON: Java Tools for Neural Networks. *Chemom. Intell. Lab. Syst.* **2002**, *61*, 167−173.
(26) The JATOON applets are available at http://www.dq.fct.unl.pt/staff/jas/jatoon.
(27) Tetko, I. V. Neural Network Studies. 4. Introduction to Associative Neural Networks. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 717−728.
(28) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: New York, 1994; p 998.

CI034228S