

## Performance Validation of Neural Network Based $^{13}\text{C}$ NMR Prediction Using a Publicly Available Data Source

K. A. Blinov,<sup>§</sup> Y. D. Smurnyy,<sup>§</sup> M. E. Elyashberg,<sup>§</sup> T. S. Churanova,<sup>§</sup> M. Kvasha,<sup>§</sup> C. Steinbeck,<sup>#</sup>  
B. A. Lefebvre,<sup>†</sup> and A. J. Williams<sup>\*,‡</sup>

Advanced Chemistry Development, Moscow Department, 6 Akademik Bakulev Street, Moscow 117513,  
Russian Federation, Advanced Chemistry Development, Inc., 110 Yonge Street, 14th floor, Toronto, Ontario,  
Canada M5C 1T4, Steinbeck Molecular Informatics, Franz-John-Strasse 10, 77855 Achern, Germany, and  
ChemZoo Inc., 904 Tamaras Circle, Wake Forest, North Carolina 27587

Received October 22, 2007

The validation of the performance of a neural network based  $^{13}\text{C}$  NMR prediction algorithm using a test set available from an open source publicly available database, NMRShiftDB, is described. The validation was performed using a version of the database containing ca. 214 000 chemical shifts as well as for two subsets of the database to compare performance when overlap with the training set is taken into account. The first subset contained ca. 93 000 chemical shifts that were absent from the ACD/CNMR DB, the “excluded shift set” used for training of the neural network and the ACD/CNMR prediction algorithm, while the second contained ca. 121 000 shifts that were present in the ACD/CNMR DB training set, the “included shift set”. This work has shown that the mean error between experimental and predicted shifts for the entire database is 1.59 ppm, while the mean deviation for the subset with included shifts is 1.47 and 1.74 ppm for excluded shifts. Since similar work has been reported online for another algorithm we compared the results with the errors determined using Robien’s CNMR Neural Network Predictor using the entire NMRShiftDB for program validation.

### 1. INTRODUCTION

Since the 1950s NMR has been widely employed to elucidate molecular structures. A combination of 1D and multinuclear NMR spectroscopy techniques makes this form of spectroscopy the definitive technique for structure elucidation. As a rule both 1D and 2D  $^1\text{H}$  and  $^{13}\text{C}$  spectra are used in combination with mass spectrometry to elucidate chemical structures. Other magnetically active nuclei, specifically  $^{15}\text{N}$  of late, are used with increased voracity as a result of improvements in both hardware and improved pulse sequences.

The analysis of  $^{13}\text{C}$  NMR spectral data is key to developing the framework of a molecule under examination, and these data can provide a large number of structural constraints when compared with other nuclei. As a result  $^{13}\text{C}$  spectra have formed the basis of many expert systems (see reviews<sup>1,2</sup> developed for the purpose of Computer-Assisted Structure Elucidation (CASE)). In these systems the program initially reveals possible molecular substructure fragments from the  $^{13}\text{C}$  NMR data then all plausible structures are generated and the most probable one selected on the basis of comparing the predicted  $^{13}\text{C}$  NMR spectra of candidates with the experimental spectrum. Since the output file of an expert system may contain hundreds and thousands of structures, a program to perform carbon chemical shift prediction must possess two properties. These are generally contradictory in nature—so the algorithm should be fast enough to deliver

predicted spectra for a large structural file in a reasonable time while maintaining high enough accuracy to provide reliable identification of the most probable structure. Spectrum prediction is also necessary to support the process of carbon chemical shift assignment and verification of the structural hypotheses. The need to create software capable of predicting  $^{13}\text{C}$  NMR spectra was realized very early on during the first steps of computerized spectroscopy development.

The first algorithms developed<sup>3–6</sup> were based on *additive rules*.<sup>7</sup> This approach allowed a chemical shift of a given carbon atom to be calculated by the means of increments characterizing different substituent patterns. Algorithms of this nature, and the software programs derived from them, are very fast, but their accuracy is unsatisfactory for the correct selection of the preferred structure in all but the simplest of cases.

Markedly higher accuracy is obtained using a *structural database* method. In this approach the environment of each carbon atom within a molecule is described by a Hierarchical Organization of Spherical Environments (HOSE) code as introduced by Bremser<sup>8</sup> and extended in other works.<sup>9–14</sup> The chemical shift is calculated by comparing the HOSE codes of the analyzed molecules with the HOSE codes of reference molecules present in the database. Our own experience with the development and implementation of HOSE code based algorithms shows<sup>15–17</sup> that this approach possesses an accuracy that, as a rule, performs admirably in allowing the program to identify the most probable structure within an answer file. Simultaneously a program equipped with appropriate functionality can provide visual elements capable

\* Corresponding author e-mail: antony.williams@chemspider.com.

† Advanced Chemistry Development, Inc.

‡ ChemZoo Inc.

§ Advanced Chemistry Development, Moscow Department.

# Steinbeck Molecular Informatics.

of explaining the details behind each calculated chemical shift value. The investigator can utilize the program to show the chemical structure and its HOSE codes whose chemical shift assignments were used during the computation of the predicted shifts. One major shortcoming of the HOSE approach is that the prediction is relatively slow, consuming some tens of seconds for complex chemical structures even on modern computers. This prevents the application of this approach to very large structural files.

$^{13}\text{C}$  chemical shift prediction programs based on *quantum-chemical calculations* have also been developed.<sup>18–20</sup> Experience has shown that they cannot be used for the purpose of prediction for large structural files since they are very time-consuming. Their accuracy, in general, can also be insufficient for the selection of the preferred structure in routine mode, though the accuracy can be refined when the program is adjusted to the shift prediction of compounds belonging to a specific class.<sup>21</sup>

In an attempt to combine both high accuracy and speed of prediction the application of artificial neural nets<sup>22</sup> (ANN) to the  $^{13}\text{C}$  spectrum prediction of different classes of organic molecules was suggested in a series of publications.<sup>23–25</sup> The ANN is efficient when the dependence of a target value on a set of parameters is either unknown or the dependence is very complex to calculate. ANNs can be taught to predict target values by using a training set of reference data. Meiler et al.<sup>13,14</sup> reported a program based on ANN algorithms allowing the prediction of  $^{13}\text{C}$  NMR spectra with average chemical shift deviations of 1.6 ppm at a calculation speed about 1000 times faster than HOSE code predictions. Therefore their work provided a good balance between accuracy and the speed of the shift prediction.

Recently an attempt was made<sup>26,27</sup> to improve on the chemical descriptor scheme suggested by Meiler and to select an appropriate ANN architecture to provide  $^{13}\text{C}$  chemical shift calculations using the commercial software program ACD/CNMR Predictor. Since the size and quality of the training set are both very important parameters, the ACD/Labs' database containing over 2 160 000  $^{13}\text{C}$  chemical shifts was utilized. To avoid overlaps with the training data 11 000 new compounds (over 150 000 chemical shifts) described in the literature during the period 2005–2006 were chosen as the test data set. The detailed description of this investigation will be provided in a separate article. Here we discuss only the important result regarding algorithmic validation: the average chemical shift deviation calculated for the test data set was less than 1.5 ppm.

The validation of the performance of NMR chemical shift prediction algorithms is a challenging problem with the primary challenge being the availability of a quality data set for validation of the prediction accuracy. If the validation data set contains a significant number of structures that are well represented in the database used as the basis of the prediction algorithms, then the validation exercise will not truly represent the challenges of prediction. The most valid test would be conducted on a validation set containing chemical structures which are very different from those contained within the training data set. Ideally, an independent party without knowledge of the structures in the training set should choose the validation set, so as to avoid any bias.

The quality of a validation database is important but difficult to prove in most cases. The ideal validation set does

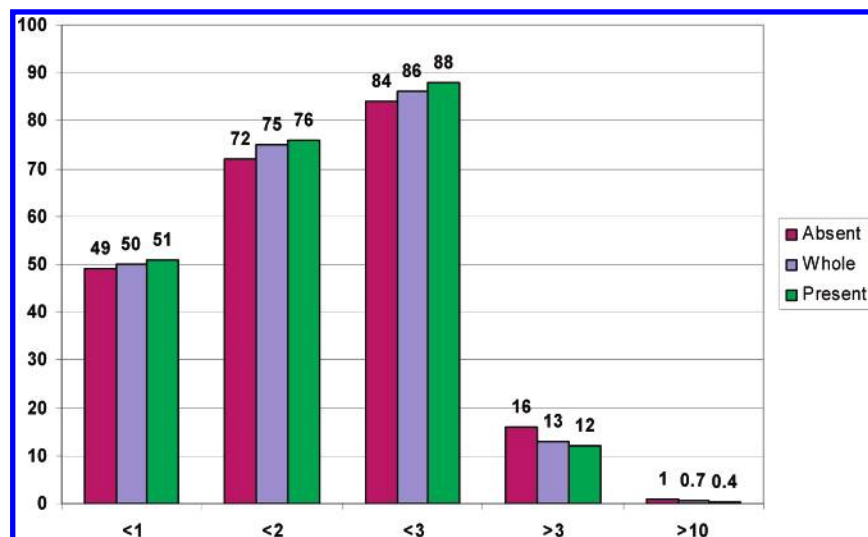
not contain any errors in assignment and covers the whole range of structural diversity available in present chemistry and in all future diversity possibilities. While this is clearly impossible to attain, large diverse data sets do exist and, while not ideal, can be used for the purpose of validation. Every large data set contains errors, but for comparisons of prediction between different algorithms this is actually irrelevant since any errors remain challenging for all algorithms.

In spite of the fact that the test data set used for validation of the improved ANN algorithm<sup>26,27</sup> had no overlap with the training set, it was of interest to determine to what degree the accuracy is dependent on the size, composition, and diversity of the structural file used as a testing set. A resource is available on the Internet that has met the above criteria of size and quality to serve as a fair and reliable validation set. This resource is a database called NMRShiftDB<sup>28–30</sup> and has been created as a collaborative effort by chemists and spectroscopists submitting data to the open access database. The current work is devoted to an analysis of the performance of the ANN based ACD/CNMR predictor using the NMR-ShiftDB database as the validation set. Due to the availability of a comparison test issued by Wolfgang Robien<sup>31</sup> we also had an opportunity to compare performance with his neural network algorithm.

## 2. NMRSHIFTDB

The NMRShiftDB is an open source collection of chemical structures and their associated NMR shift assignments. The database is generated as a result of contributions by the public and has been described in detail elsewhere.<sup>28,29</sup> Currently, the database contains 19 958 structures with 214 136 assigned carbon chemical shifts. Data sets entered by contributors are sent to registered reviewers for evaluation, who are presented the newly entered spectrum together with a shift prediction and a color-coded table of deviations. A significant part of NMRShiftDB, however, was initially assembled from in-house databases from collaborating institutions and has been entered unchecked. This called for external checks of the data based on independent databases and resources, which have now been carried out as described in this paper. Based on a cursory examination of the structural diversity within the database these data represent a statistically relevant set to use in an evaluation of predictive accuracy and is the first large data set available from an independent source which we could use for this purpose.

Robien has already published an analysis of performance of his neural network predictions.<sup>31</sup> This review provides an evaluation of the NMR prediction algorithms he has developed over many years. These algorithms have been the basis of a number of software products including a commercially available product, NMRPredict.<sup>31</sup> Robien's analysis focused on the quality of the database in terms of the presence of a number of outliers but gave no specific review of the quality of the data set and focused only on the problem assignments. A later comment suggested that about 0.3% of the data may be in error, a low number for the purpose of the work reported here.



**Figure 1.** The percentage of chemical shift deviations depending on different ranges of values as calculated for the three data sets presented in Table 1.

**Table 1.** General Results of ACD/CNMR Neural Network Predictor Validation

|                         | shift count | av error (ppm) | std error (ppm) | max error (ppm) | calculation method |          |          |          |           |
|-------------------------|-------------|----------------|-----------------|-----------------|--------------------|----------|----------|----------|-----------|
|                         |             |                |                 |                 | % <1 ppm           | % <2 ppm | % <3 ppm | % >3 ppm | % >10 ppm |
| 1. whole data set       | 214136      | 1.59           | 2.76            | 153.53          | 50                 | 75       | 86       | 13       | 0.7       |
| 2. absent in data file  | 92927       | 1.74           | 3.22            | 133.19          | 49                 | 72       | 84       | 16       | 1         |
| 3. present in data file | 121209      | 1.47           | 2.35            | 153.53          | 51                 | 76       | 88       | 12       | 0.4       |

### 3. PERFORMANCE VALIDATION OF ACD/LABS NEURAL NETWORK CNMR PREDICTION

As already discussed the NMRShiftDB Web site offers visitors the opportunity to download a data file containing all of the structures and chemical shifts that compose the database. This file was downloaded, and the structures and shifts were imported into ACD/Labs' format.

As a first step, an analysis of the degree of overlap between the structures in the training set within the ACD/CNMR Predictor and the validation set made up of NMRShiftDB was undertaken. The presence or absence of a chemical shift in the NMRShiftDB was determined in the following manner. For a specific carbon atom existing in a structure contained within the NMRShiftDB, the HOSE code was determined, and then the atom with the same HOSE code was searched in the ACD/CNMR DB. If such an atom was identified, then the corresponding chemical shift was deemed to be present in both databases and therefore excluded. Otherwise, the shift was included in the data set for which the given chemical shift was absent from NMRShiftDB.

It was determined that 57% of the carbon chemical shifts in the NMRShiftDB were already contained within the ACD/Labs database. The NMRShiftDB database was stripped of replicate chemical shifts used as the basis of the prediction algorithms in ACD/CNMR Predictor.

The results of algorithm validation using the NMRShiftDB test data set are shown in Table 1.

It was revealed that from the total number of chemical shifts collected in the NMRShiftDB (214 136) that 92 927 shifts were not contained within the ACD/CNMR database, and consequently their HOSE codes were new for the NN ACD/CNMR Predictor. At the same time 121 209 chemical shifts made up the overlap of the two databases. The average, standard, and maximum errors are displayed for the whole

database and two subsets as well as the percentages of chemical shifts predicted with errors of  $d < 1$  ppm,  $d < 2$  ppm,  $d < 3$  ppm,  $d > 3$  ppm, and  $d > 10$  ppm for the entire database and both subsets.

Table 1 shows that the mean error calculated for the entire data set is 1.59 ppm (rms = 2.76 ppm). This value can vary by  $\pm 9\%$  depending on whether data set 2 or 3 is used for the purpose of algorithm validation. Since the NMRShiftDB library is composed of compounds analyzed by chemists working in varied areas of organic chemistry, we believe that it forms a representative set for general chemistry. Assuming this to be true then validation of our  $^{13}\text{C}$  chemical shift prediction method using a database of ca. 214 000 chemical shifts leads to an average deviation which is reliable within  $\pm 10\%$ .

The percentage of chemical shift deviations depending on the different ranges of the values calculated for the three data sets presented in Table 1 are shown in Figure 1.

About 50% of chemical shifts were calculated with mean errors  $< 1$  ppm in all three cases. This observation allows us to conclude that independent of the presence or absence of the predicted chemical shifts in the test database half of all of the predicted chemical shifts are calculated with an error of  $< 1$  ppm. Errors  $< 2$  and  $< 3$  ppm encompass 72–76 and 84–88% of all calculated shifts, respectively. Chemical shifts predicted with relatively low accuracy ( $d > 3$  ppm) make up only 12–16% of the entire shift number. This number can correspond to errors in the database, and some have already been identified by Robien<sup>31</sup> and in this work. However, this distribution also correlates with the distribution of problems solved with the aid of the Structure Elucidator software with chemical shift deviations calculated for genuine structures.<sup>17</sup> Thus, the deviations calculated by ACD/CNMR Predictor based on HOSE codes<sup>10</sup> were  $> 3$  ppm for 18% of

**Table 2.** Whole Data Set<sup>a</sup>

|                    | atom types |       |       |       |             |              |             |            |             |
|--------------------|------------|-------|-------|-------|-------------|--------------|-------------|------------|-------------|
|                    | C          | CH    | CH2   | CH3   | alkene<br>C | alkene<br>CH | alkyne<br>C | aroma<br>C | aroma<br>CH |
| number of shifts   | 6108       | 17238 | 38296 | 28952 | 20192       | 11143        | 2693        | 38379      | 51135       |
| mean deviation     | 1.79       | 2.01  | 1.55  | 1.17  | 2.02        | 2.04         | 2.40        | 1.74       | 1.26        |
| standard deviation | 3.31       | 3.32  | 2.98  | 2.26  | 3.66        | 3.24         | 4.45        | 2.60       | 1.90        |

<sup>a</sup> Deviations calculated for different atom types.**Table 3.** Absent in Data File<sup>a</sup>

|                     | atom types |      |       |       |             |              |             |            |             |
|---------------------|------------|------|-------|-------|-------------|--------------|-------------|------------|-------------|
|                     | C          | CH   | CH2   | CH3   | alkene<br>C | alkene<br>CH | alkyne<br>C | aroma<br>C | aroma<br>CH |
| number of shifts:   | 3116       | 9792 | 18205 | 12570 | 8628        | 4396         | 932         | 15289      | 19999       |
| mean deviations     | 1.93       | 2.16 | 1.71  | 1.30  | 2.23        | 2.27         | 2.51        | 1.96       | 1.30        |
| standard deviations | 3.69       | 3.80 | 3.66  | 2.44  | 4.15        | 3.89         | 4.50        | 3.00       | 2.09        |

<sup>a</sup> Atom types.**Table 4.** Present in Data File<sup>a</sup>

|                     | atom types |      |       |       |             |              |             |            |             |
|---------------------|------------|------|-------|-------|-------------|--------------|-------------|------------|-------------|
|                     | C          | CH   | CH2   | CH3   | alkene<br>C | alkene<br>CH | alkyne<br>C | aroma<br>C | aroma<br>CH |
| number of shifts:   | 2992       | 7446 | 20091 | 16382 | 11564       | 6747         | 1761        | 23090      | 31136       |
| mean deviations     | 1.64       | 1.82 | 1.41  | 1.07  | 1.86        | 1.90         | 2.35        | 1.58       | 1.24        |
| standard deviations | 2.87       | 2.56 | 2.19  | 2.12  | 3.24        | 2.74         | 4.42        | 2.30       | 1.78        |

<sup>a</sup> Atom types.

all of the problems solved with Structure Elucidator, and, in each case, a newly identified natural product was elucidated. The percentage of shifts predicted with errors >10 ppm varies from 0.4% ("present") to 1% ("absent") and are mainly associated with the presence in NMRShiftDB of some outliers. These outliers can be from poor prediction versus experimental or experimental data in error as a result of an incorrect structure representation or misassignment. These errors are distributed as follows:  $d > 10$  ppm (1040; 0.5%),  $d > 25$  ppm (141; 0.07%), and  $d > 50$  ppm (31; 0.01%). Examination of the structure giving rise to the maximum error of ~150 ppm provides evidence that NMRShiftDB either contains some erroneously drawn structures or poorly assigned chemical shifts.<sup>31</sup> This is unavoidable in a large database of this nature. Examples of structures for which obviously erroneous chemical shift assignments were detected (marked by red) as a result of the prediction accuracy validation are shown in Figure 2.

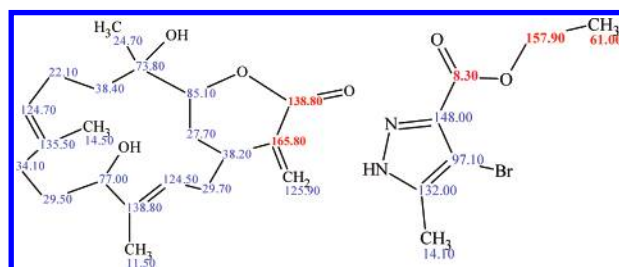
When the shifts with difference >25 ppm were removed from the entire database and the two subsets, the average errors decreased only slightly: from 1.59 to 1.56 ppm for the entire DB, from 1.74 to 1.70 ppm for the "absent" subset, and from 1.47 to 1.46 ppm for the "present" subset. Consequently, the conclusion regarding the predictive ability of the ACD/CNMR NN Predictor remains.

The data presented in Table 1 and Figure 1 describe the average values characterizing the results of the validation study. The accuracy of <sup>13</sup>C NMR spectrum prediction is known to depend on chemical classes to which a given carbon atom belongs, on the number of representatives of a given class in the training sets, and on the influence of stereochemical factors. We have therefore compared the average prediction accuracy for quaternary carbons, methine, ethyl, and methyl groups in aliphatic substructures as well

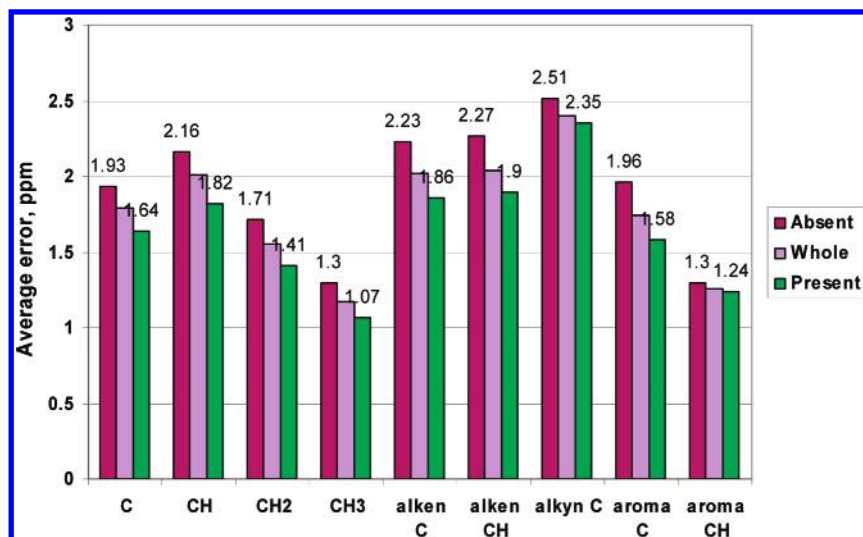
as the accuracy for alkenes, alkynes, and aromatic compounds. In so doing we have examined how the errors of chemical shift prediction are associated with the presence or absence of a given shift in the validation set. The results of the calculations performed with the whole database and both subsets are presented in Tables 2–4. In these tables the mean and standard deviations are shown along with the numbers of shifts predicted for each type of carbon atoms.

To assist in the analysis of the results they are graphically represented in Figure 3.

A striking observation is that the highest and almost equal accuracy is achieved for the methyl group ( $d = 1.30$ – $1.07$  ppm) and the aromatic protonated carbon atoms ( $d = 1.30$ – $1.24$  ppm), and the errors associated with the aromatic carbons are almost the same for the entire database and both of its subsets. The next in the series is the –CH<sub>2</sub>– group for which the error varies between 1.41 and 1.71 ppm. At first glance one can expect that the accuracy of chemical shift prediction for >CH– groups should be higher than for quaternary carbons since the latter have four non-hydrogen neighbors whose influence on chemical shifts is complex and difficult to take into account. We however observe the

**Figure 2.** Examples of structures for which obviously erroneous chemical shift assignments were detected (marked by red) as a result of the prediction accuracy validation.





**Figure 3.** The comparison of chemical shift prediction accuracy for different types of atoms depending on their assignment to one of the three testing sets.

**Table 5.** Validation of the ACD/CNMR NN Predictor and Robien's Algorithms on the Entire Data Set

| program         | shift count | average deviation (ppm) | standard deviation (ppm) | outliers (ppm difference) |             |            |
|-----------------|-------------|-------------------------|--------------------------|---------------------------|-------------|------------|
|                 |             |                         |                          | > 10 ppm                  | > 25 ppm    | > 50 ppm   |
| ACD/CNMR v10.05 | 214 136     | <b>1.59</b>             | 2.76                     | 1040 (0.4%)               | 141 (0.07%) | 31 (0.01%) |
| Robien          | 209 412     | <b>2.22</b>             | N/A                      | N/A                       | 194 (0.09%) | 56 (0.03%) |

reverse: 1.64–1.93 ppm for quaternary carbons and 1.82–2.16 ppm for >CH– (see Figure 3). The accuracy of the chemical shift prediction for =CH and quaternary =C groups in olefins is almost the same if it is considered separately within the entire set and both subsets. The error varies from 1.86 to 2.27 ppm when going from the “present” subset to “absent”. The relatively large errors are accounted for by the difficulties associated with an attempt to allow for stereochemical factors playing an important role in the chemical shift prediction of double bond carbons. The lowest accuracy of shift prediction ( $d = 2.35$ –2.51 ppm) was observed for alkynes, which can be explained by a very reduced data set for such compounds in the training set.

#### 4. COMPARISON WITH RESULTS OF PREVIOUS VALIDATION WORK PERFORMED USING NMRSHIFTDB

As mentioned earlier the entire NMRShiftDB was recently used by Robien<sup>31</sup> to validate the performance of his <sup>13</sup>C NMR shift prediction algorithms based on artificial neural nets. The validation provided an average deviation of 2.22 ppm. Since these calculations were performed only with the entire DB, we cannot comment on the overlap between his training set and the test set. We can only compare the results for the entire database (see Table 5). It should be noted that the data set tested by ourselves was a later data set than that examined by Robien and contained more data points.

The comparison shows that the average deviation obtained by ACD/CNMR Predictor (1.59 ppm) is 40% lower than that obtained by Robien, outperforming by a significant margin.

It should be noted that only 203 284 unique carbon centers were represented in the NMRShiftDB, but some had multiple assignments. All redundancy was included in case there was disagreement between the assignments. Therefore over

214 000 assignments were considered in our calculations. As commented on earlier this is more than the number used by Robien (209 412 chemical shifts). Only a small number of the errors identified by Robien in his analysis had been corrected, but the majority remained unchanged. The presence of outliers was almost the same in both investigations. The Web posting of Robien<sup>31</sup> did not provide a measure of standard deviation or the number of chemical shift predictions that were more than 10 ppm from their experimental value, and this is the reason for the absence of this parameter in the table. Robien quoted an average deviation of 2.19 ppm after correction of some errors, but, for comparison purposes, we have used the 2.22 ppm value with no corrections since no corrections were made to the data set examined with the ACD/CNMR Predictor.

#### 5. CONCLUSION

The NMRShiftDB is an excellent resource for the purpose of evaluating chemical shift prediction accuracy as evidenced by this work and the previous work of Robien. As identified by Robien initially, and later in this work, there are certainly outliers in the data set requiring review and correction. Our previous work has shown that the literature itself contains about 8% errors in the form of misassignments, transcription errors, and incorrect structures. The obvious errors in NMRShiftDB are certainly below this level, and this is a testament to the value of this resource. The NMRShiftDB data set is large and structurally diverse and continues to grow as scientists contribute. Both this study and the one performed by Robien also demonstrate the virtues of open access to primary research data, since they would not have been possible had NMRShiftDB been a closed resource.

Despite a large overlap between the NMRShiftDB and the ACD/Labs carbon NMR database, a statistically relevant validation set of ca. 93 000 chemicals shifts was extracted

from the NMRShiftDB and used to test the algorithms. The data presented here show that the ACD/Labs prediction algorithms have an average deviation of less than 1.8 ppm on the validation set ("absent" subset of entire NMRShiftDB) and significantly outperforms the algorithms of Robien presented in his review. The algorithm defining the ACD\CNMR neural network predictor discussed in detail in ref 27.

## REFERENCES AND NOTES

- (1) Munk, M. E. Computer-Based Structure Determination: Then and Now. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 997–1009.
- (2) Elyashberg, M. E.; Williams, A. J.; Martin G. E. Computer-Assisted Structure Verification and Elucidation Tools in NMR-Based Structure Elucidation. *Prog. NMR Spectrosc.* **2007**, doi:10.1016/j.pnmrs.2007.04.003.
- (3) Grant, D. M.; Paul, E. G. Carbon-13 Magnetic Resonance. II. Chemical Shift Data For The Alkanes. *J. Am. Chem. Soc.* **1964**, *86*, 2984–2990.
- (4) Clerc, J. T.; Sommerauer, H. A. A Minicomputer Program Based On Additivity Rules For The Estimation Of  $^{13}\text{C}$  NMR Chemical Shifts. *Anal. Chim. Acta* **1977**, *95*, 33–40.
- (5) Fürst, A.; Pretsch, E. A Computer Program for the Prediction of  $^{13}\text{C}$  NMR Chemical Shifts of Organic Compounds. *Anal. Chim. Acta* **1990**, *229*, 17–25.
- (6) Chen, L.; Robien, W. OPSI: A Universal Method For Prediction Of Carbon-13 NMR Spectra Based On Optimized Additivity Models. *Anal. Chem.* **1993**, *65*, 2282–2287.
- (7) Pretsch, E.; Clerc, T.; Seibl, J.; Simon, W. *Tables of Spectral Data for Structure Determination of Organic Compounds*; Springer-Verlag: Berlin, 1989.
- (8) Bremser, W. HOSE—A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355–365.
- (9) Chen, L.; Robien, W. The CSEARCH-NMR Data Base Approach To Solve Frequent Questions Concerning Substituent Effects On  $^{13}\text{C}$  NMR Chemical Shifts. *Chemom. Intell. Lab. Syst.* **1993**, *19*, 217–223.
- (10) *ACD/CNMR Predictor, version 11.0*; Advanced Chemistry Development: Toronto, Canada, 2007.
- (11) *Specinfo*; Chemical Concepts GmbH: D-69442 Weinheim, Germany.
- (12) Bremser, W. Expectation Ranges Of  $^{13}\text{C}$  NMR Chemical Shifts. *Magn. Reson. Chem.* **1985**, *23*, 271–275.
- (13) Meiler, J.; Meusinger, R.; Will, M. Fast Determination of  $^{13}\text{C}$ -NMR Chemical Shifts Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1169–1176.
- (14) Meiler, J.; Maier, W.; Will, M.; Meusinger, R. Using Neural Networks for  $^{13}\text{C}$  NMR Chemical Shift Prediction—Comparison with Traditional Methods. *J. Magn. Reson.* **2002**, *157*, 242–252.
- (15) Blinov, K. A.; Carlson, D.; Elyashberg, M. E.; Martin G. E.; Martirosian, E. R.; Molodtsov, S. G.; Williams, A. J. Computer Assisted Structure Elucidation of Natural Products with Limited Data: Application of the StrucEluc System. *Magn. Reson. Chem.* **2003**, *41*, 359–3724.
- (16) Elyashberg, M. E.; Blinov, K. A.; Molodtsov, S. G.; Williams, A. J.; Martin, G. E. *Structure Elucidator: A Versatile Expert System for Molecular Structure Elucidation from 1D and 2D NMR Data and Molecular Fragments*. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 771–792.
- (17) Elyashberg, M. E.; Blinov, A.; K.; Williams, A. J.; Molodtsov, S. G.; Martin, G. Are Deterministic Expert Systems For Computer Assisted Structure Elucidation Obsolete? *J. Chem. Inf. Model.* **2006**, *46*, 1643–1656.
- (18) *Gaussian 03, Revision C.02*; Gaussian Inc.: Wallingford, CT, 2003.
- (19) *COSMOS (Computer Simulation of Molecular Structures), version 5*; COSMOS Software GbR: Jena, Germany, 2007.
- (20) Bagno, A.; Saielli G. Computational NMR Spectroscopy: Reversing the information Flow. *Theor. Chem. Acc.* **2007**, *117*, 603–619.
- (21) Rychnovsky, S. D. Predicting NMR Spectra by Computational Methods: Structure Revision of Hexacyclinol. *Org. Lett.* **2006**, *8*, 2895–2898.
- (22) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH: Weinheim, 1993.
- (23) Kvasnicka, V.; Sklenak, S.; Pospichal, J. Application Of Recurrent Neural Network In Chemistry: Prediction And Classification Of  $^{13}\text{C}$  NMR Chemical Shifts In A Series Of Monosubstituted Benzenes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 742–747.
- (24) Mitchell, B. E.; Jurs, P. C. Computer Assisted Simulation of  $^{13}\text{C}$  Nuclear Magnetic Spectra of Monosaccharides. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 58–64.
- (25) Ivanciuc, O.; Rabine, J.-P.; Cabrol-Bass, D.; Panaye, A.; Doucet, J. P.  $^{13}\text{C}$  NMR Chemical Shift Prediction of the  $\text{sp}^3$  Carbon Atoms in the  $\alpha$ -Position Relative to the Double Bond in Acyclic Alkenes. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 587–598.
- (26) Smurnyy, Y. D.; Blinov, K. A.; Elyashberg, M. E.; Lefebvre, B. A.; Williams, A. J. NMR Chemical Shift Prediction by Atomic Increment Based Algorithms. In *Experimental Nuclear Magnetic Resonance Conference*, 48th ENC Conference Proceedings, Daytona Beach, U.S.A., April 22–28, 2007, PH-150.
- (27) Smurnyy, Y. D.; Blinov, K. A.; Churanova, T. S.; Elyashberg, M. E.; Williams, A. J. Towards More Reliable  $^{13}\text{C}$  and  $^1\text{H}$  Chemical Shift Prediction: A Systematic Comparison of Neural Network and Least Squares Regression Based Approaches. *J. Chem. Inf. Model.* **2008**, *48*, 128–134.
- (28) Steinbeck, C.; Krause, S.; Kuhn, S. NMRShiftDB—Constructing A Chemical Information System With Open Source Components. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1733–1739.
- (29) Steinbeck, C.; Kuhn, S. NMRShiftDB — Compound Identification And Structure Elucidation Support Through a Free Community-Built Web Database. *Phytochemistry* **2004**, *65*, 2711–2717.
- (30) The NMRShift Database. <http://www.nmrshiftdb.org> (accessed September 23, 2007).
- (31) A Quality Check of the NMRShiftDB using the CSEARCH Algorithms. [http://nmrpredict.orc.univie.ac.at/csearchlite/enjoy\\_its\\_free.html](http://nmrpredict.orc.univie.ac.at/csearchlite/enjoy_its_free.html) (accessed June 6, 2007).

CI700363R