Revue Roumaine de Chimie, **1995**, 40 (10), 1093–1101

This paper is dedicated to the memory of Professor ŞERBAN SOLACOLU (1905–1980)

ARTIFICIAL NEURAL NETWORKS APPLICATIONS. PART 6[1]

USE OF NON-BONDED VAN DER WAALS AND ELECTROSTATIC INTRAMOLECULAR ENERGIES IN THE ESTIMATION OF ^{13}C–NMR CHEMICAL SHIFTS IN SATURATED HYDROCARBONS

OVIDIU IVANCIUC

Department of Organic Chemistry, Faculty of Chemical Technology, University "Politehnica" Bucharest, Splaiul Independenţei 313, 77206 Bucharest, Roumania

Received March 7, 1995

The ^{13}C-NMR chemical shifts of alkanes and cycloalkanes were estimated with multi-layer feed forward neural networks, using as structural descriptors the degree of the carbon atom, the non-bonded van der Waals intramolecular energy, and the local electrostatic energy. The neural network Quantitative Structure Property Relationship (QSPR) model provides better results than the multilinear regression one. The predictive ability of the neural networks was tested with leave-one-out method. Neural networks with a bell-shape hidden activation function or with linear output activation function were tested, with good results.

INTRODUCTION

Artificial Neural Networks (ANN) are algorithmic systems derived from a simplified concept of the brain. [2,3] In a neural network, a number of computational units, called artificial neurons, are interconnected into a net-like structure. A network is constructed with three or more layers of neurons: input neurons, output neurons, and one or more layers of intermediate elements called the hidden neurons.

Recent progress in ANN, mainly the multilayer feed-forward (MLF) neural networks trained with the backpropagation algorithm, offers new computational models with a high impact in chemistry. [4,5] Some important ANN applications in organic magnetic resonance are represented by the prediction of ^{13}C–NMR chemical shifts in alkanes (only for secondary carbon atoms), [6] in keto-steroids, [7] in monosubstituted benzenes, [8,9] and in a large set of variously branched alkanes. [10] Also, neural networks were used to predict the phosphorus NMR shifts, [11] for classifying cross peaks in two-dimensional NMR spectra, [12,13] and for the identification of 2D proton NMR-antiphase cross peaks. [14]

The goal of the present paper is to compare the performance of a multilinear regression (MLR) model and of a neural model in estimating the ^{13}C—NMR chemical shifts of alkanes and cycloalkanes. The structural descriptores are the degree of the carbon atom, the non-bonded van der Waals intramolecular energy, E_{vdW}, and the local electrostatic energy, E_q.

RESULTS AND DISCUSSION

A recently proposed empirical analysis of the ^{13}C—NMR chemical shifts for numerous classes of unsubstituted hydrocarbons of varied degrees of strain used two atomic parameters: the local non-bonded van der Waals, E_{vdW}, and the electrostatic, E_q, intramolecular interaction energies.[15] The E_{vdW} term for an atom i equals the van der Waals interaction, computed by the MM2 procedure, between all pairs of atoms i and j separated by

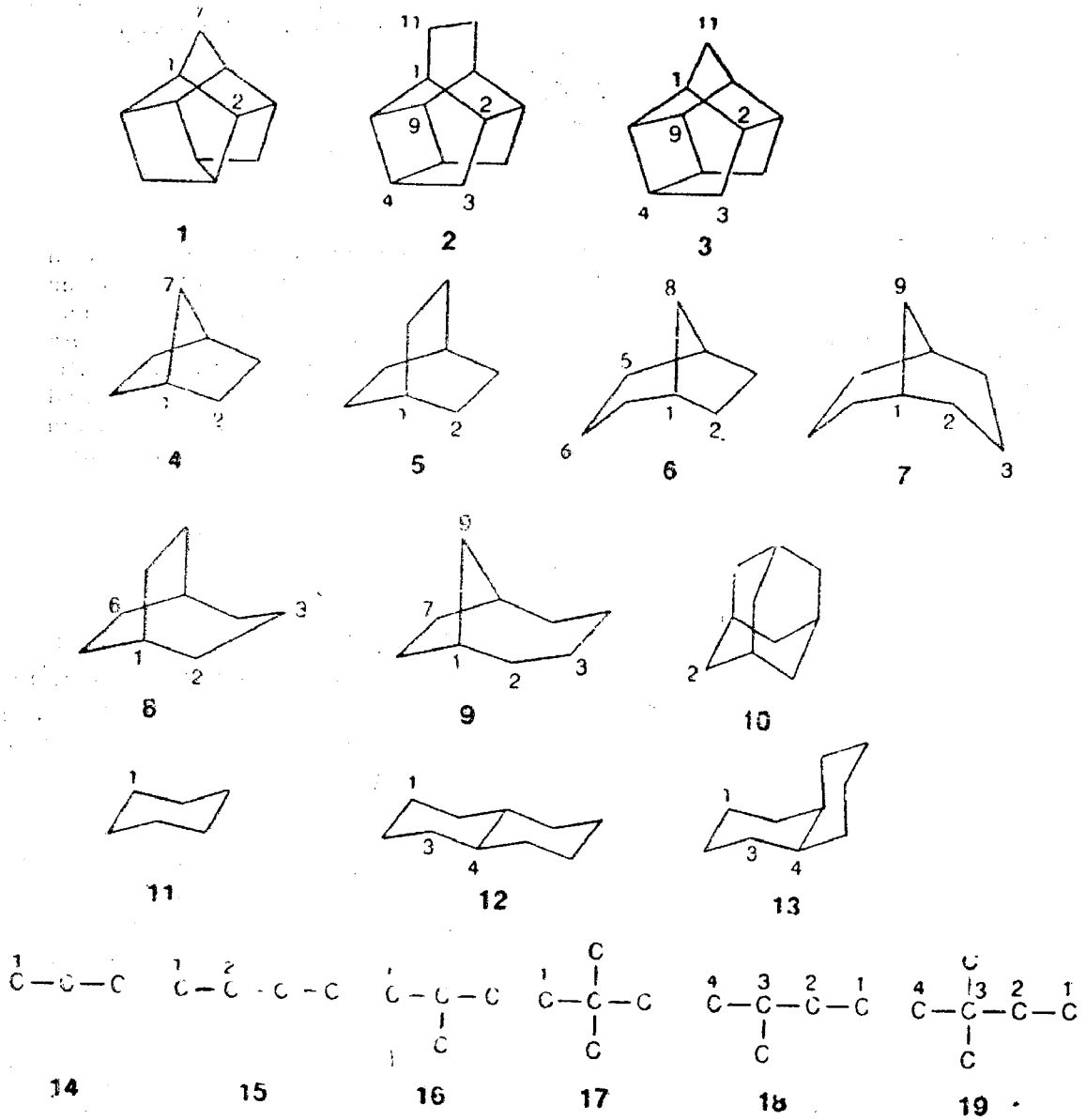

Fig. 1. — Structure of the saturated hydrocarbons whose ^{13}C-NMR chemical shifts are reported in Table 1.

more than two bonds. The net atomic charge q_i was computed with Gaussian-76 at the STO-3G level, and the electrostatic interaction energy for atom i was computed with the formula: $E_{qi} = q_i \Sigma q_j/R_{ij}$, in kcal/mol, where the summation goes for all atoms j situated at more than two bonds from i. The relationship was successful when applied to the estimation of the ^{13}C-NMR chemical shifts of primary, secondary and tertiary carbon atoms of numerous classes of unsubstituted saturated hydrocarbons of varied degrees of strain presented in Fig. 1. Table 1 lists the energies E_{vdW} and E_q in kcal/mol, the ^{13}C-NMR chemical shifts δ in ppm, and the degree Deg for the carbon atoms in molecules **1—19** used to develop the models in this papers. All data are taken from ref. 15. The degree of a carbon atom is 1 for a primary, 2 for a secondary, and 3 for a tertiary atom. Using the data for the 59 carbon atoms in Table 1, we have computed the following MLR model:

$$\delta_i = -2.416(\pm 1.067) + 12.84(\pm 5.67)\mathrm{Deg}_i - \quad (1)$$

$$- 2.900(\pm 1.282)E_{vdWi} - 3{,}698(\pm 1.634)E_{qi}$$

$$n = 59 \qquad r = 0.874 \qquad s = 4.153 \qquad mres = 3.09$$

where n is the number of compounds used in the correlation, r is the correlation coefficient, s is the standard deviation, and mres is the mean residual, computed with the formula: $mres = \Sigma|\delta_{calc} - \delta_{exp}|/n$. The standard error of estimation of each coefficient at the 95% confidence level is given in parentheses. The partial correlation coefficients are $r(Deg) = 0.654$, $r(E_{vdW}) = 0.200$, and $r(E_q) = -0.176$, leading to the conclusion that no single parameter can estimate the δ values. The correlation coefficient of eq. (1) is low, and this relationship has only a qualitative value, a fact indicated by the high values of s and mres. The collinearity between the three variables is low, as indicated by the intercorrelation matrix:

	Deg_i	E_{vdWi}	E_{qi}
Deg_i	1.000	0.238	0.485
E_{vdWi}		1.000	−0.141

The low value for the correlation coefficient in the MLR model has two possible explanations: (a) the dependence between the three atomic parameters and δ is strongly nonlinear; (b) not all important structural parameters which influence the ^{13}C-NMR chemical shifts were identified and used in the MLR model. If the low performance of the MLR model is due to the factor (a), the ANN model will give better results, otherwise, in the case (b), the neural model will fail to give a good fit for the ^{13}C-NMR chemical shift.

The neural networks used in the present study are three-layer MLF networks, with three input units representing the three parameters in eq. (1)· (Deg, E_{vdW}, and E_q), and one output unit representing the values of δ from Table 1. The training was done with the backpropagation algorithm,[2] the input and output data scalings were set between −0.9 and 0.9 (with the

Table 1

Degrees (Deg), local van de Waals energies (E_{vdW}, kcal/mol), charge interaction energies (E_q, kcal/mol), and ^{13}C-NMR chemical shifts (ppm) for the carbon atoms in molecules **1—19**

Molecule	Atom i	Deg_i	E_{vdWi}	E_{qi}	δ_i
14	1	1	0.1665	−2.6038	15.60
15	1	1	0.1503	−1.4097	13.20
16	1	1	0.3824	−5.1646	24.30
17	1	1	0.6501	−7.6783	31.50
18	1	1	0.0832	−0.0361	11.50
18	4	1	0.4024	−3.8795	22.00
19	1	1	0.0135	1.2990	8.70
19	4	1	0.6843	−6.3716	28.90
1	7	2	−0.1648	−1.4914	33.20
2	3	2	1.3266	−1.1345	27.51
2	11	2	0.0959	−0.7372	19.39
3	3	2	0.7622	−1.1386	27.00
3	11	2	−0.3067	−1.4811	33.70
4	2	2	0.5766	−2.0979	29.80
4	7	2	0.1828	−3.9461	38.40
5	2	2	0.7198	−1.7470	26.10
6	2	2	0.4160	−1.9048	28.90
6	5	2	0.6897	−2.4802	32.80
6	6	2	0.5714	−0.1196	19.10
6	8	2	0.4292	−3.5188	39.70
7	2	2	0.2812	−2.3559	31.60
7	3	2	0.1521	−0.0611	22.50
7	9	2	0.6998	−3.1186	35.10
8	2	2	0.2774	−2.3072	35.70
8	3	2	0.3509	−0.0206	22.40
8	6	2	0.4802	−1.6222	25.90
8	2	2	0.3803	−2.3713	35.90
9	2	2	0.1257	−0.7518	25.60
9	3	2	0.0528	−1.8806	33.10
9	7	2	0.1878	−3.3044	35.50
9	9	2	0.8613	−3.3925	37.80
10	2	2	0.3324	−1.5381	27.70
11	1	2	0.2338	−1.0505	27.30
12	1	2	0.4522	−1.7258	34.70
12	3	2	0.1488	−0.9944	24.60
13	1	2	0.3553	−1.6580	29.80
13	3	2	0.1487	−1.3790	25.00
15	2	2	0.2997	−2.7591	31.80
18	2	2	0.5048	−4.1239	36.70
19	2	2	0.1641	−0.9225	47.70
1	1	3	1.7504	−2.0157	41.50
1	2	3	1.5756	−0.9517	35.21
2	1	3	1.1250	−1.0240	37.87
2	2	3	0.4327	−0.8357	37.47
2	4	3	0.6663	−1.2876	38.68
2	9	3	0.0259	−0.9553	46.70
3	1	3	1.0113	−1.1761	42.30
3	2	3	0.4427	−0.8268	36.30
3	4	3	0.6033	−1.5506	43.70
3	9	3	−0.2137	−0.6570	36.40
4	1	3	1.2796	−0.6990	24.00
5	1	3			

Table 1 (continued)

1	2	3	4	5	6
6	1	3	0.1669	−0.5690	35.20
7	1	3	0.2754	−0.4650	27.90
8	1	3	0.6090	−0.6539	29.00
9	1	3	−0.1750	−0.5338	37.40
10	1	3	0.5748	−0.1103	28.50
12	4	3	0.6243	−0.9280	44.20
13	4	3	0.4842	−0.9378	36.90
18	3	3	0.0711	−0.2515	29.90

exception of the networks with linear output activation function), the initial weights scaling between −0.1 and 0.1, the learning rate 0.01, and the momentum 0.8. The usual activation function has a sigmoid shape, but we have investigated the use of a bell-shape activation function in the hidden layer, and of a linear function for the output layer. The bell-shape function was computed with the formula: $Act\ (z) = 1/(1 + z)^2$. Another activation function used in this study was the hyperbolic tangent, tanh. The training was done by randomly presenting to the network the values of the parameters for the 59 types of carbon atoms from Table 1, until the correlation coefficient between experimental and computed δ improved with less than 10^{-4} in 100 epochs.

Because we need some statistical indices in order to compare the performance of the ANN model with the results of the MLR equation (1), we have used the standard deviation s, the correlation coefficent r, and the mean residual mres, of the linear correlation between δ_{exp} (the experimental ^{13}C-NMR chemical shift) and δ_{ANN} (the ^{13}C-NMR chemical shift computed by the ANN model):

$$\delta_{exp} = A + B\ \delta_{ANN}$$

We have to point that the topology of the ANN determines its performances. The number of neurons in the hidden layer was selected on the basis of empirical trials, in which ANN with different numbers of hidden neurons were trained until convergence was obtained. A neural network (denoted by NN1) with two hidden neurons was trained for 3100 epochs, and provided the following results in estimation δ:

$$\delta_{exp} = 0.32 + 0.97\ \delta_{NN1} \qquad (2)$$

$$n = 59 \quad r = 0.923 \quad s = 3.229 \quad mres = 2.595$$

The value of the correlation coefficient in eq. (2) shows an important improvement when compared with that of the MLR model in eq. (1). In order to investigate the predictive character of the ANN model, we have used the leave-one-out (LOO) cross-validation method. In the LOO technique, an untrained network is first created, then the first pattern is deleted from the training set of patterns and the network is trained with the remaining patterns. When the learning process is finished, the network predicts the output value for the pattern which was eliminated from the learning set.

The pattern is then put back in the set and the next one is deleted and the training process is repeated, starting with the untrained network. In this cross-validation method, each pattern serves as a prediction pattern once and as a training pattern all the other times. The LOO predictions of the neural network NN1 are well correlated with the experimental values:

$$\delta_{exp} = 1.05 + 0.95\ \delta_{NN1\ LOO} \tag{3}$$

$$n = 59 \qquad r = 0.889 \qquad s = 3.836 \qquad mres = 3.101$$

An increasing of the size of the hidden layer to three neurons provides a better neural model:

$$\delta_{exp} = 0.01 + 0.99\ \delta_{NN2} \tag{4}$$

$$n = 59 \qquad r = 0.950 \qquad s = 2.610 \qquad mres = 2.062$$

The LOO predictions of the NN2 network are of the same statistical quality as those for the network NN1:

$$\delta_{exp} = 1.48 + 0.95\ \delta_{NN2\ LOO} \tag{5}$$

$$n = 59 \qquad r = 0.888 \qquad s = 3.855 \qquad mres = 3.075$$

With four hidden neurons, the ANN model with tanh activation functions offers a fairly good model, which could not be improved by increasing the number of hidden neurons:

$$\delta_{exp} = -0.44 + 1.00\ \delta_{NN3} \tag{6}$$

$$n = 59 \qquad r = 0.958 \qquad s = 2.401 \qquad mres = 1.810$$

The prediction power of the NN3 network is almost unchanged when compared with that of the previous two networks:

$$\delta_{exp} = 3.35 + 0.88\ \delta_{NN3\ LOO} \tag{7}$$

$$n = 59 \qquad r = 0.887 \qquad s = 3.867 \qquad mres = 3.158$$

If we compare the statistical indices of the MLR model represented by eq. (1) with the results of the ANN model in eq. (2), (4), and (6), it is clear that neural network outperforms regression analysis and provides superior estimation of the ^{13}C-NMR chemical shifts using the three atomic parameters.

It is very important to take into consideration that feedforward neural networks are universal approximators, and are capable of arbitrarily accurate approximation of arbitrary functions, when sufficiently many hidden units are available. [16–18] The potential of chance correlations in ANN was investigated for classification [19] and for continous output networks [20] using random data as input and output patterns. Because the random generated data do not adequately represent the type of patterns normally encountered in a quantitative structure-property relationship study, we used two methods to detect the presence of chance effects in the results of the ANN model.

The first method used a simulated learning set obtained by retaining the δ values of the output patterns assigned to the corresponding atoms. and randomly generating the input data. A neural network with four hidden neurons was trained for 2000 epochs using this simulated training set. If the fit obtained with the real data is consistently better than that with the random input data, the correlation obtained with the real data can confidently be said to be not due to chance factors. The procedure of generating simulated training sets was repeated ten times. Because the mean standard deviation for the ten runs was 6.443, the mean r was 0.617, and the maximum correlation coefficient was 0.771, we consider that the results reported for ANN are not due to chance effects.

Because the simulation of learning in ANN using random input data give no idea of the inhomogeneous distribution of the values of parameters, the second method uses an artificial training set obtained from the real one by randomly assigning a real δ value to a carbon atom not necessarily possessing this value, and retaining the values of the input patterns. A neural network with four hidden neurons was trained for 2000 epochs with 10 such random training sets. For the ten runs, the mean standard deviation was 7.650, the mean correlation coefficient 0.368, and the maximum r 0.628. Taking into consideration the fact that we found that the statistical indices in the runs using random input data and randomly reassigned output data are considerably lower than those obtained when the real training set of data is used, we are confident that a valid relationship has indeed been established between the three atomic descriptors and ^{13}C-NMR chemical shifts.

In our search for a better ANN model, we have investigated the performances of a neural network provided with a linear activation function in the output layer. The network NN4, with four hidden neurons, output scaling between -5 and 5, and a linear output function, was trained for 10000 epochs, giving a slightly better estimation of the δ values than the network NN3:

$$\delta_{exp} = -0.99 + 1.01\,\delta_{NN4} \qquad (8)$$

$$n = 59 \qquad r = 0.961 \qquad s = 2.320 \qquad mres = 1.801$$

The LOO cross-validation indicates that the NN4 network offers the best predictions when compared with that of the previous networks, indicating that a linear output activation function gives to the network better generalization characteristics for the patterns considered in our investigation:

$$\delta_{exp} = 0.51 + 0.97\,\delta_{NN4\,LOO} \qquad (9)$$

$$n = 59 \qquad r = 0.918 \qquad s = 3.326 \qquad mres = 2.578$$

While the usual activation function has the shape of a sigmoid, [2,3] we have tested a bell-shaped activation function $Act(z)$, for the neurons in the hidden layer, given by the equation: $Act(z) = 1/(1 + z^2)$. The network NN5, with four hidden neurons, a bell-shaped hidden function, and the tanh output function was trained for 30000 epochs, and provided the best estimation for the ^{13}C-NMR chemical shifts:

$$\delta_{exp} = -0.12 + 1.00\,\delta_{NN5} \qquad (10)$$

$$n = 59 \qquad r = 0.961 \qquad s = 2.316 \qquad mres = 1.804$$

While the estimations of the NN5 network are of good statistical quality, the predictions computed by the LOO cross-validation method show that the bell hidden activation function has a low generalization potential:

$$\delta_{exp} = 5.54 + 0.81\ \delta_{NN5\ LOO} \qquad (11)$$

$$n = 59 \qquad r = 0.820 \qquad s = 4.798 \qquad mres = 3.784$$

The networks NN4 (with a linear output function) and NN5 (with the bell-shaped hidden function) provide a small improvement when compared with the network NN3 (with tanh activation functions), but we consider that they are good alternatives to the usual sigmoid activation functions. On the other hand, the fact that for the set of patterns considered in this paper the linear output function gives the best LOO predictions, while the bell hidden activation function provides poor predictions, deserves further investigations before a general conclusion will be obtained.

The results presented in this study show that higher estimations of the ^{13}C-NMR chemical shifts are obtained with the ANN model, which outperforms the MLR model, and we can draw the conclusion that the dependence between the three atomic parameters (Deg, E_{vdW}, and E_q) and δ is nonlinear. But the maximum correlation coefficient for the estimations of the ANN model is 0.960, and the standard deviation is much higher than the experimental error in determining the ^{13}C-NMR chemical shifts. An explanation for this fact is that not all important structural parameters which influence the ^{13}C-NMR chemical shifts were identified and used in the model. Work is in progress in order to identify other structural factors which determine the ^{13}C-NMR chemical shifts.

The results obtained in the present study show that the use of ANN has some important advantages in developing quantitative structure-property relationships models, because the mathematical form of the relationship between the input and output data does not need to be provided, and ANN are able to represent nonlinear relationships between input and output patterns.

ACKNOWLEDGEMENT. We thank the Ministry of Research and Technology for financial support of this research under Grants 572B TA4 and TB11.

REFERENCES

1 Previous part: O. Ivanciuc, *Rev. Roum. Chim.*, in press.
2 D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, **1986**, *323*, 533–536.
3 P. D. Wasserman, *Neural Computing*, Van Nostrand Reinhold, New York, 1989.
4 J. Gasteiger and J. Zupan, *Angew. Chem. Int. Ed., Engl.*, **1993**, *32*, 503–527.
5 J. R. M. Smits, W. J. Melssen, L. M. C. Buydens and G. Kateman, *Chemom. Intell. Lab. Syst.*, **1994**, *22*, 165–189.
6 V. Kvasnička, *J. Math. Chem.*, **1991**, *6*, 63–76.
7 L. S. Anker and P. C. Jurs, *Anal. Chem.*, **1992**, *64*, 1157–1164.
8 V. Kvasnička, S. Sklenák and J. Pospichal, *J. Chem. Inf. Comput. Sci.*, **1992**, *32*, 742–747.
9 V. Kvasnička, S. Sklenák and J. Pospichal, *J. Mol. Struct. (Theochem)*, **1992**, *277*, 87–107.

[10] J. P. Doucet, A. Panaye, E. Feuilleaubois and P. Ladd, *J. Chem. Inf. Comput. Sci.*, **1993,** *33*, 320—324.
[11] G. M. J. West. *J. Chem. Inf. Comput. Sci.*, **1993,** *33*, 577—589.
[12] S. A. Corne, A. P. Johnson and J. Fisher, *J. Magn. Reson.*, **1992,** *100*, 256—266.
[13] S. A. Corne, J. Fishe , A. P. Johnson and W. R. Newell, *Anal. Chim. Acta*, **1993,** *278*, 149—158.
[14] M. Kjaer and F. M. Poulsen, *J. Magn. Reson.*, **1991,** *94*, 659—663.
[15] R. Pachter and P. L. Wessels, *Magn. Reson. Chem.*, **1989,** *27*, 277—282.
[16] G. Cybenko, *Math. Control Signals Syst.*, **1989,** *2*, 303—314.
[17] K. Funahashi, *Neural Networks*, **1989,** *2*, 183—192.
[18] K. Hornik, M. Stinchcombe and H. White, *Neural Networks*, **1989** *2*, 359—366.
[19] D. T. Manallack and D. J. Livingstone, *Med. Chem. Res.*, **1992,** *2*, 181—190.
[20] D. J. Livingstone and D. T. Manallack, *J. Med. Chem.*, **1993,** *36*, 1295—1297.