JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 10, 612–622 (1997)

APPLICATION OF NEURAL NETWORKS TO STRUCTURE–SANDALWOOD ODOUR RELATIONSHIPS

DRISS ZAKARYA,[1]* DRISS CHERQAOUI,[2] M'HAMED ESSEFFAR,[2] DIDIER VILLEMIN[3] AND JEAN-MICHEL CENSE[4]

[1] *Département de Chimie, Faculté des Sciences et Techniques, Mohammadia, Morocco*
[2] *Département de Chimie, Faculté des Sciences Semlalia, BP S 15, Université Cadi Ayyad, Marrakech, Morocco*
[3] *Ecole Nationale Supérieure d'Ingénieurs de Caen (ENSI de Caen), ISMRA, URA 480 CNRS, 6 boulevard Maréchal Juin, 14050 Caen Cedex, France*
[4] *Ecole Nationale Supérieure de Chimie de Paris, 11 rue P. et M. Curie, 75005 Paris, France*

Neural networks have proved to be particularly successful in their ability to identify non-linear relationships. This paper shows that a three-layer back-propagation neural network is able to learn the relationship between the sandalwood odour and molecular structures of 85 organic compounds belonging to acyclic, cyclohexyl, norbornyl, campholenyl and decalin derivatives. Four steric and three electronic parameters were used to describe each molecular structure. Odour was coded by a binary variable. The neural network was used to classify the compounds into two groups and to predict their odours (sandalwood or non-sandalwood). The results obtained were compared with those given by discriminant analysis, and found to be better. The most important descriptors were revealed on the basis of correlation analysis. © 1997 John Wiley & Sons, Ltd.

J. Phys. Org. Chem. **10**, 612–622 (1997) No. of Figures: 2 No. of Tables: 2 No. of References: 48
Keywords: neural networks; structure–odour relationships; sandalwood
Received 19 February 1996; revised 10 February 1997; accepted 27 February 1997

INTRODUCTION

The molecular mechanisms involved in olfaction are not well defined. However, several elements have been discovered and may be useful for understanding the mechanisms of the sense of smell. In addition to the important investigations of Amoore,[1] Beets,[2] Ohloff[3] and Chastrette and Zakarya,[4] many studies have been carried out and have shown the existence of structure–odour relationships (e.g. musky,[5–7] sandalwood,[8] camphoraceous,[9] bitter almond[10, 11] and amber[12]).

Sandalwood is an appreciated fragrance corresponding to well defined chemical structures.[8] This was explained by the empirical rules of Brunke and Klein[13] and Naipawer *et al*.[14] Buchbauer *et al*.[15] have shown that the molecular surface is closely related to sandalwood character, but Chastrette *et al*.[8] have found that only a part of the molecule is responsible for the interaction between sandalwood odorants and olfactive receptor sites. They also defined a santalophore superpattern by superimposition of several sandalwood leader molecules.

In the present study, we tried to establish structure–sandalwood odour relationships with neural networks using descriptors known to be responsible for the sandalwood odour.

Neural networks[16] (NNs) are artificial systems simulating the function of the brain where very high number of information-processing neurons are interconnected. They can handle problems involving imprecise or 'noisy' data as well as problems that are highly non-linear and complex. NNs can identify and learn correlative patterns between sets of input data and corresponding target values. An NN must be trained by being repeatedly fed input data together with their corresponding target outputs. After training, the NN has been initiated to recognize the relationship between input and output data and creates an internal model as a governing data process. The NN can then use this internal model to make predictions for new inputs.

The application of NNs to solving problems in chemistry[17, 18] is a recent field of research. NNs have been applied to the investigation of quantitative structure–activity relationships (QSAR),[19–22] structure–musky odour relationships,[23, 24] estimation of physical properties,[25–27] prediction of chemical reactivity,[28, 29] identification of proton NMR spectra,[30] interpretation of IR spectra,[31, 32] prediction of ^{13}C chemical shifts,[33] classification of mass spectra[34] and determination of protein structure.[35, 36]

* *Correspondence to:* D. Zakarya.

 CCC 0894–3230/97/080612–11 $17.50

EXPERIMENTAL

Compounds and descriptors used. Eighty-five compounds[37] were studied (39 sandalwood and 46 non-sandalwood odorants): 7 acyclic, 19 cyclohexyl, 21 norbornyl, 32 campholenyl and 6 decalin derivatives (Figure 1). The chosen chemical structures represent the main categories of compounds developing sandalwood fragrance.

It is believed that the sandalwood odour might be closely related to the size, shape and functionality of a molecule.[8, 13, 14] These three factors can be described by the molecular surface area (S), the molecular volume (V), the ovality (O) of the molecule, the AM1 dipole moment (D), the molecular weight (W), the square root of the sum of the squared AM1 charges on oxygen atoms (Q_0) and the ionization potential ($I_p = -$HOMO), which is a parameter taking the molecular electronic state into account. Q_0 was recently used by Bodor *et al*.[38] and found to be correlated to log P.

Starting geometries of molecules were generated using HyperChem[39] and fully optimized geometries were obtained with the AM1 Hamiltonian[40] with AMPAC on a Silicon Graphics computer. In this study, the possibility of interaction of less stable conformers was not examined.

Geometric parameters were computed by numerical integration techniques similar to those described by Gavezzotti.[41] For the molecular volume, the molecule was immersed in a regular three-dimensional grid bound by the three-dimensional extent of the molecule and every grid point was tested for inclusion into any atom of the molecule. The grid spacing was 0·1. The molecular surface area was calculated from a regular grid of points disposed on the surface of each atom; 4558 points were used for hydrogen atoms and 10 270 points for the other atoms. The ovality estimation used was that given by Bodor *et al*.[26]:

$$O = S/(4\pi K), \qquad K = 3V/(4\pi)^{2/3} \tag{1}$$

Dipole moment, charge, molecular weight and ionization potential were read directly from the AMPAC output.

Neural network. All the feed-forward NNs used in this paper are three-layer networks with seven units in the input layer, a variable number of hidden neurons and one unit in the output layer. A bias term was added to the input and hidden layers. Figure 2 shows an example of the architecture of such an NN. Each neuron of the input layer is fully interconnected with each neuron of the hidden layer, which in turn is fully interconnected with the output neuron. There is no connection between the neurons within a layer nor any direct connection between those of the input and output layers. Input and output data are normalized between 0·1 and 0·9. The sigmoidal transfer function used for NN is given by the equation

$$O_i = [1 + \exp(-\sum W_{ij} O_j)]^{-1} \tag{2}$$

where O_i and O_j are the outputs of neuron i and j, respectively, and W_{ij} is the weight connecting neuron i to neuron j.

The output of NN describing the odour is coded 1 if the molecule is sandalwood and 0 otherwise. In this paper, for NN output values below 0·4 or above 0·6, the prediction was considered as correct for both non-sandalwood and sandalwood compounds. When the NN output values were between 0·4 and 0·6, the molecule classification was said to be incorrect by this network. We could have chosen 0·5 as a threshold value between sandalwood and non-sandalwood compounds, but to be more accurate, we decided to choose two limits.

The connection weights between the neurons were initially assigned random values uniformly distributed between $-0{\cdot}5$ and $+0{\cdot}5$ and no momentum was added. The back-propagation (BP) algorithm was used to adjust those weights. This algorithm has been described previously[42] with a simple example of application and details of this algorithm are given elsewhere.[42, 43, 44] The learning rate was initially set to 1 and was gradually decreased until the error function could no longer be minimized.

All calculations of NNs were performed on an 80486 personal computer running at 33 MHz using our program written in C language.

RESULTS AND DISCUSSION

Three different aspects were considered: classification, prediction and interpretation of the relationships obtained.

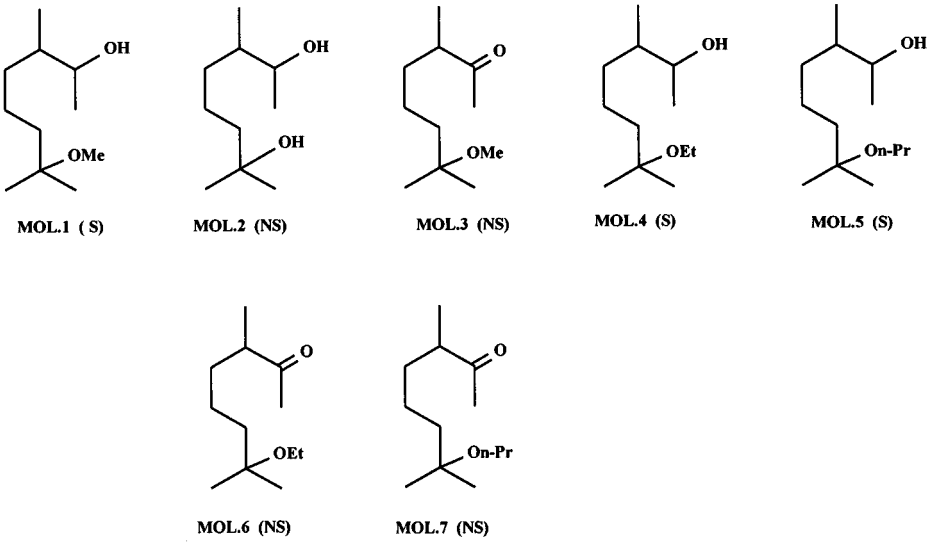
Classification

Descriptors are a crucial part of any attempt to apply NNs to classification problems, data analysis, etc. Input data must provide the greatest amount of information possible on the molecules. In order to ensure that the seven descriptors retained (S, V, O, D, W, Q_0, I_p) were adequate, seven architectures were tried ($7-x-1$; x=4, 5, 6, 7, 8, 9, 10). Each architecture was trained with five different initial random sets of weights (35 architectures were tested with all the compounds). After the training phase, the classification ability of NNs was applied to the same 85 compounds. Out of the whole database, only one molecule (No. 75) was wrongly classified by all the architectures tried. The chemical structure of this molecule is closely similar to those of the other norbornyl derivatives, except for the fact that it includes three double bonds, which is probably the cause of the absence of sandalwood odour.

Prediction

In this preliminary study, we were unable to draw conclusions about the ability of NNs to establish a satisfactory relationship between the seven descriptors and the sandalwood odour. Since its predictive ability is one of the most important attributes of an NN, this ability was used to classify the molecules according to their odour. However, one of the major problems arising when trying to get the NN

Acyclic derivatives



Cyclohexyl derivatives

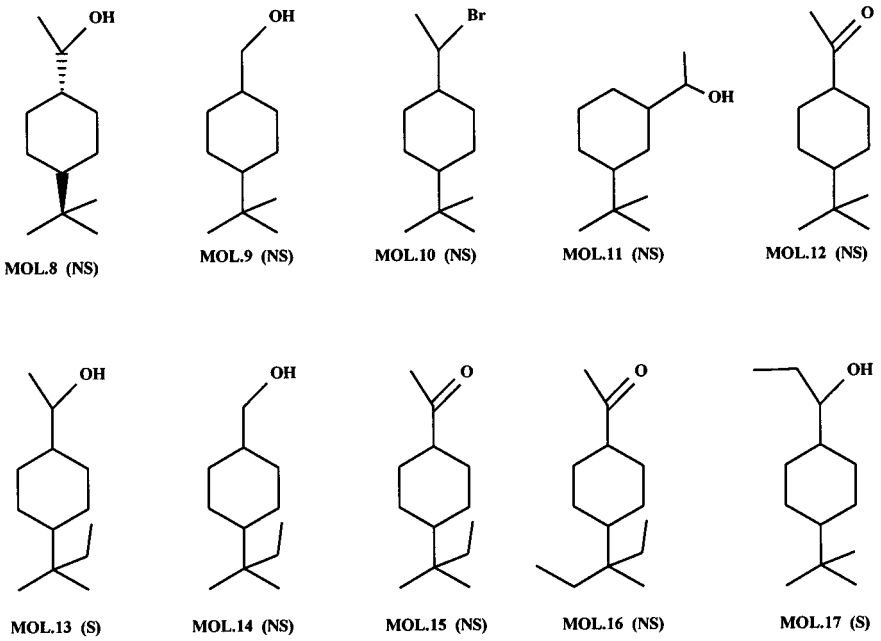
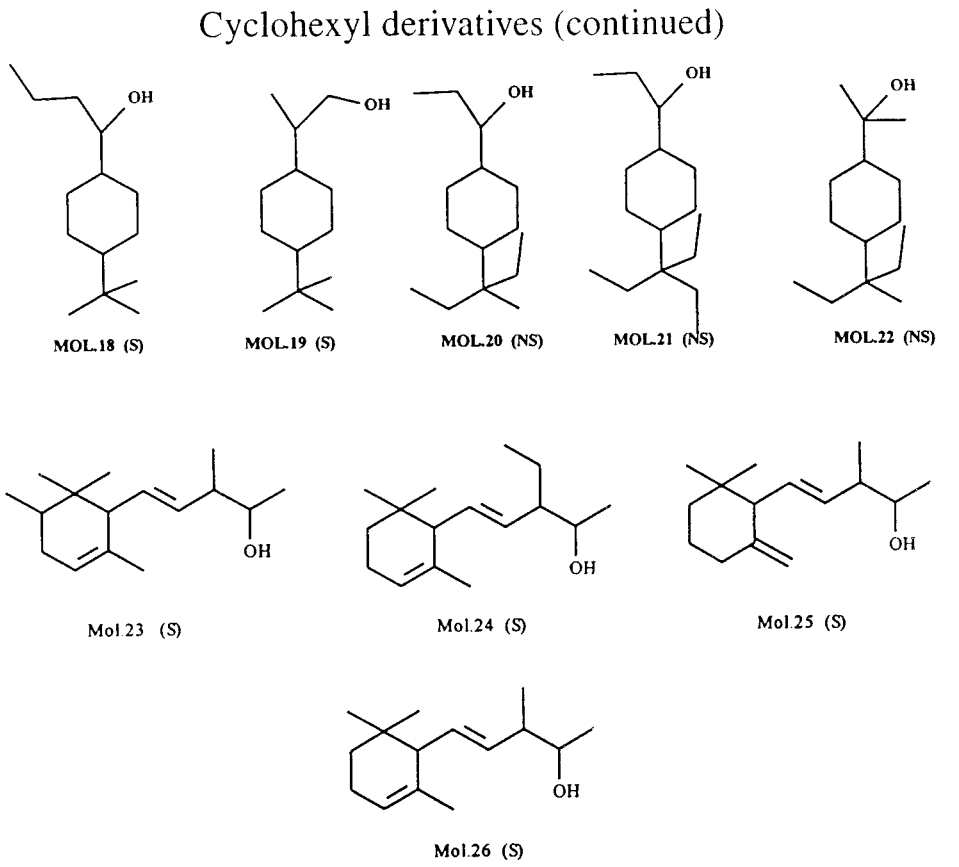


Figure 1. Structures of the compounds studied



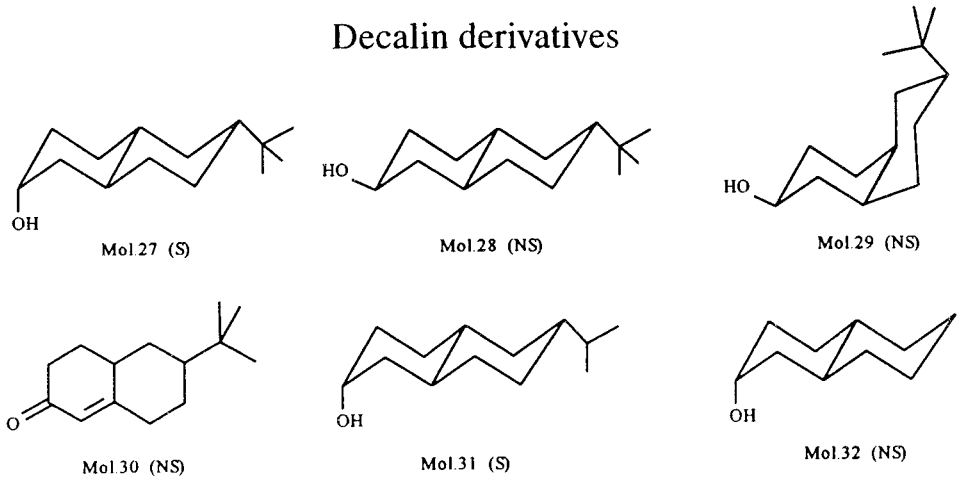


Figure 1. Continued

Campholenyl derivatives

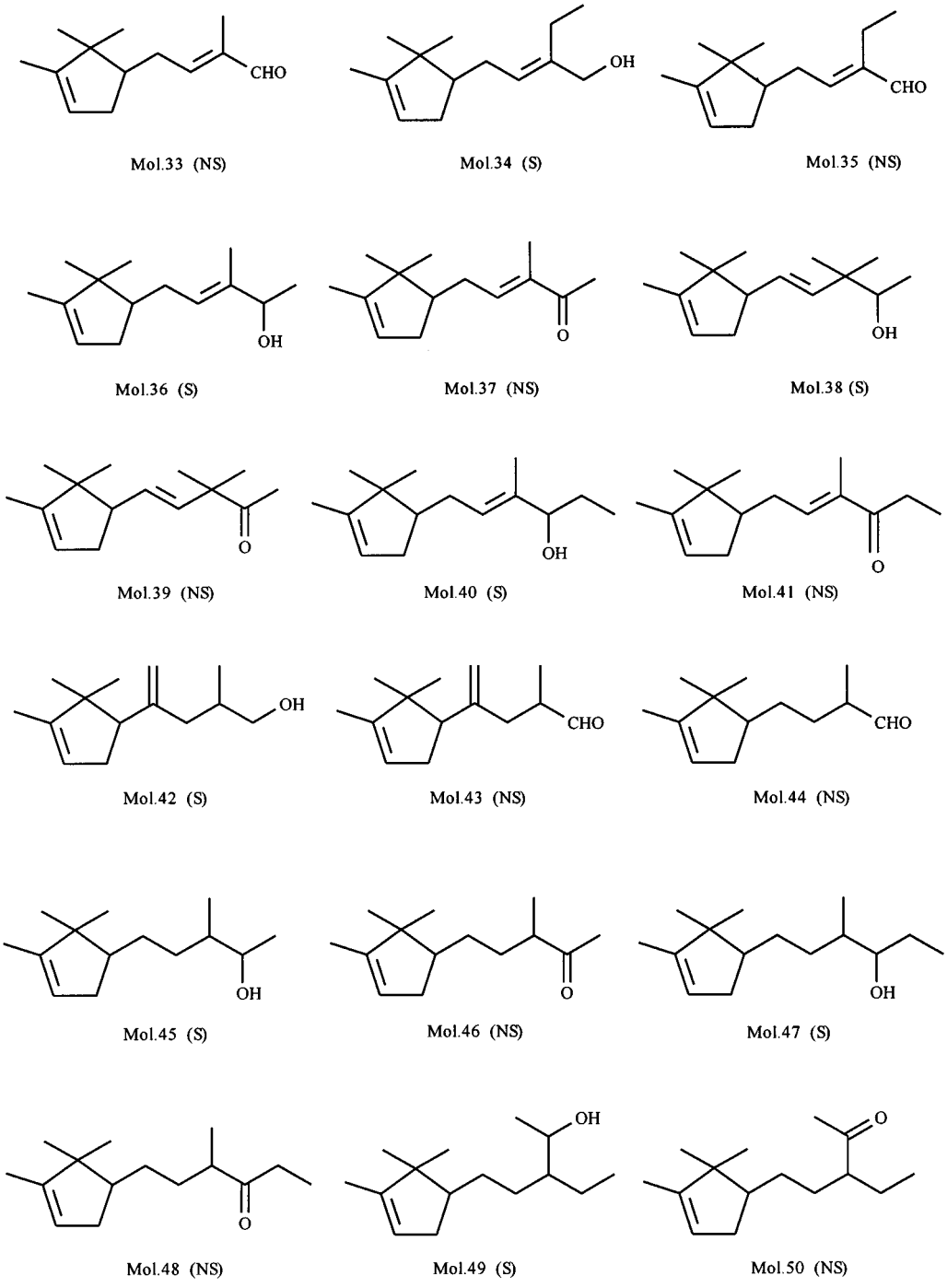
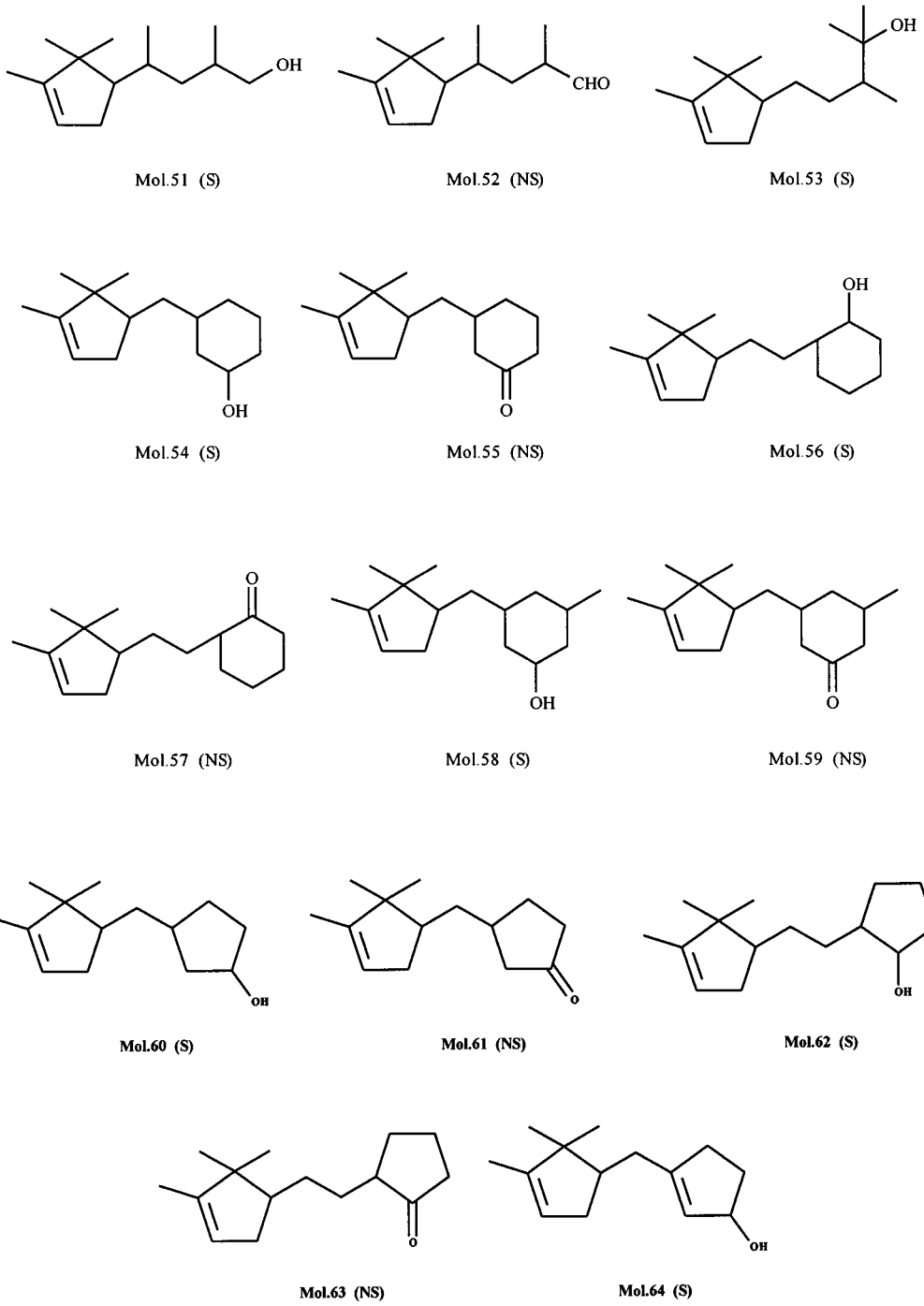


Figure 1. Continued

Campholenyl derivatives (continued)



Mol.51 (S) Mol.52 (NS) Mol.53 (S)

Mol.54 (S) Mol.55 (NS) Mol.56 (S)

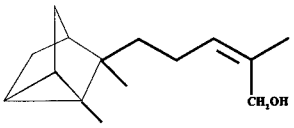
Mol.57 (NS) Mol.58 (S) Mol.59 (NS)

Mol.60 (S) Mol.61 (NS) Mol.62 (S)

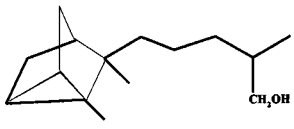
Mol.63 (NS) Mol.64 (S)

Figure 1. Continued

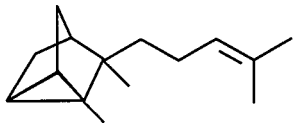
Norbornyl derivatives



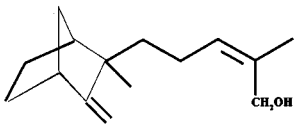
Mol.65 (S)



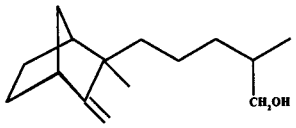
Mol.66 (S)



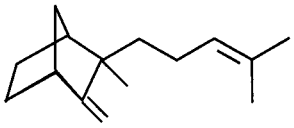
Mol.67 (NS)



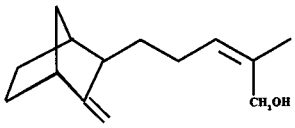
Mol.68 (S)



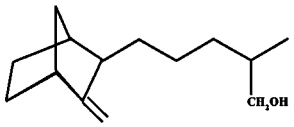
Mol.69 (S)



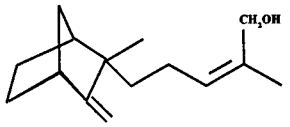
Mol.70 (NS)



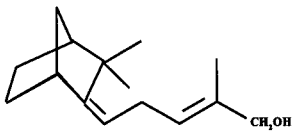
Mol.71 (S)



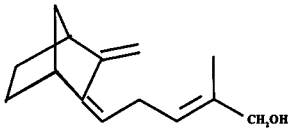
Mol.72 (S)



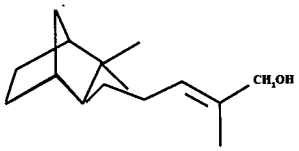
Mol.73 (S)



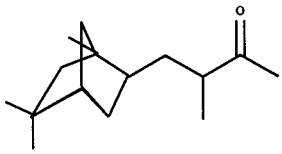
Mol.74 (S)



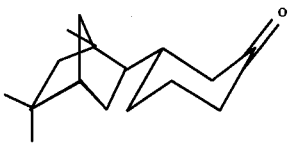
Mol.75 (NS)



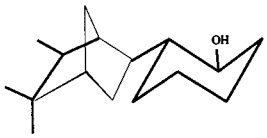
Mol.76 (S)



Mol.77 (NS)



Mol.78 (NS)



Mol.79 (NS)

Figure 1. Continued

Norbornyl derivatives (continued)

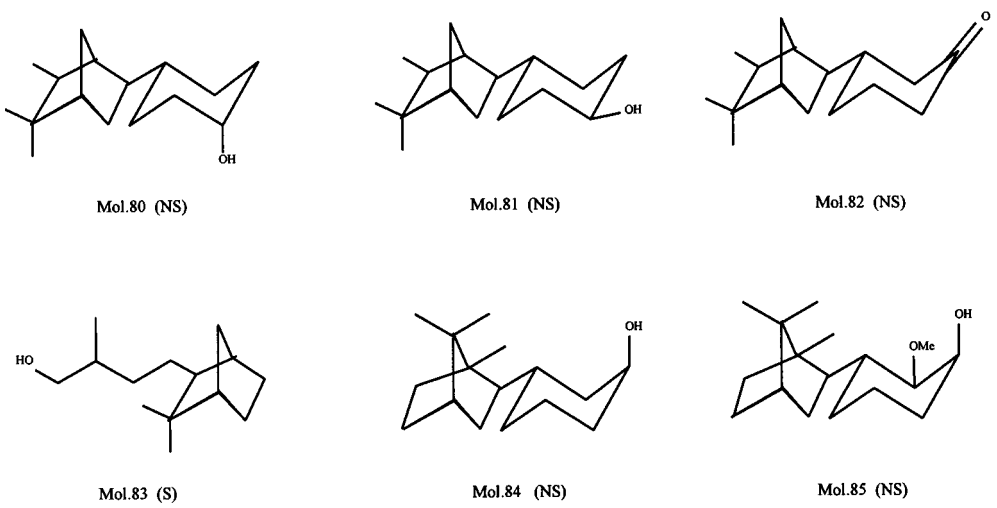


Figure 1. Continued

to be good at prediction was the choice of its architecture. In a BP NN, one of the main issues is to determine how many hidden layers and how many neurons in these layers could be used for any particular problem. The number of hidden layers necessary for a successful QSAR depended on the complexity of the problem to be solved. However, for most of the applications of NNs to chemistry, one hidden layer seems to be sufficient. The number of hidden neurons determines the number of adjustable parameters (connection strengths) of the NN model. If too few hidden neurons are used, then the learning process will be hindered. In the same way, if too many hidden neurons are included, the NN will have a tendency to memorize the training data and therefore the net's generalizing capacity will decrease. In order to determine the best architecture, the simplest approach is to proceed for a random partitioning of the available data into a training and a test set. The former set generally contains more samples than the latter. The best network in those trained is then determined by the minimum in the test set error. However, the performance of an NN depends on many factors, including the degree of correspondence between the training and test sets, which is not taken into account when partitioning the data randomly. A more sophisticated approach, used to optimize the number of hidden neurons, is the cross-validation. In this procedure, one compound is removed from the data set and the network is trained with the remaining compounds and used to predict the discarded compound. The process is repeated for each compound in the data set. The NN with the smallest error on test set is then selected. In this work, cross-validation was used to evaluate the predictive ability of the NN. Seven architectures were tried ($7-x-1$; x=4, 5, 6, 7, 8, 9, 10) and the maximum number of iterations was set to 1000 but sufficient convergence was usually obtained after 200 iterations. The computation time varied depending on the architecture of the NN and also on the size of the data set. In this study, the average training time for each run was about 3 h. Cross-validation was also used to assess the predictive power of the discriminant analysis (DA). DA was applied to the same data set and to the same seven molecular descriptors. The results obtained are given in Table 1. They are satisfactory and showed once again that the seven

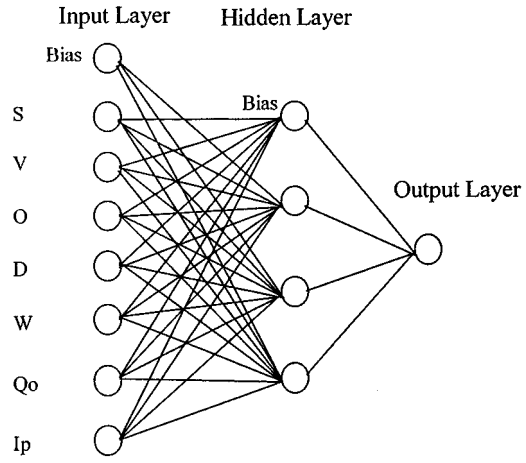


Figure 2. Schematic representation of a three-layer neural network. The configuration shown is $7-3-1$ (the bias unit is not included in the unit count)

Table 1. Comparison of predictive ability for neural networks and discriminant analysis using the cross-validation procedure

NN configuration	Number of well classified molecules	% of correct classification
7 − 4 − 1	76	89·4
7 − 5 − 1	76	89·4
7 − 6 − 1	75	88·2
7 − 7 − 1	77	90·6
7 − 8 − 1	79	92·9
7 − 9 − 1	75	88·2
7 − 10 − 1	75	88·2
DA	73	85·9

molecular features are very useful descriptors for the compounds studied. Among all the architectures of NNs, the best one (7 − 8 − 1) correctly predicted the odour of 79 molecules out of the 85 (92.9%).

In the best architecture, the ratio (ρ) value is in agreement with the values recommended by Andrea and Kalayeh,[45] Chastrette *et al*.[46] and Zupan and Gasteiger.[47]

The six remaining compounds were distributed as following: three of them (Nos 2, 66 and 75) appeared in the wrong category, two (Nos 2 and 75) were predicted as being a sandalwood odorant when they were not, and one (No. 66) as a non-sandalwood odorant when it was. The last molecule is probably weak sandalwood owing to the absence of a double bond in the side-chain[8, 13, 14] in comparison with similar sandalwood structures. We do not have an explanation for molecule No. 2 and we consider that the model established is probably not highly efficient for such a class of molecules. The other three molecules (Nos 27, 31 and 56) had a tendency to be well predicted but still remained outside the limits fixed beforehand (their outputs were between 0·4 and 0·6). They are described as sandalwood and their calculated odour values were 0·54, 0·52 and 0·58, respectively.

The results given by all the architectures tried were more satisfactory than those of the DA method. Thus, we found that the NN gave excellent use of the information included in the given data compared with a conventional method. In DA, the relationship between sandalwood odour and the seven descriptors is expressed by a linear combination. In contrast, one crucial aspect of the predictive performance of an NN used to solve classification problems is its non-linear power. Note that in the present work, DA gave a rate of prediction of 86%, indicating that the function mapped by the network is not so far from linear.

Interpretation

Very good correlations were obtained which allowed the prediction of the sandalwood character of new molecules. Moreover, it is interesting to evaluate the importance (contribution) of each descriptor used. The contribution of a descriptor was estimated starting from the trained 7 − 8 − 1 configuration network. The descriptor under study was removed from the set of seven descriptors together with its corresponding weights. Then the network (6 − 8 − 1) calculated the output of each molecule as usual. The mean deviation Δm between the observed sandalwood odour (0 or 1) and the predicted one for each class of compounds was calculated. Table 2 gives the results obtained. For example, the column under the heading *S* gave Δm when the descriptor *S* was removed. It can be noted that in most cases the values of Δm are high, which once again shows the relevance of the descriptors chosen.

The established model may be a useful tool for a preliminary screening of hypothetical sandalwood molecules. Although it is said that only part of the sandalwood molecules interact with receptor sites, the size of those molecules must be included in the range 12–17 carbon atoms.[13, 14, 48] This is taken into account by the molecular weight *W* and the molar volume *V*. The parameters *S*, *V* and *O* account for the global size, the shape of the molecule and the steric interactions that could play a role in the establishment of the structure–odour relationships.

Descriptors related to the electronic structure of the molecule seem to be very important in the discrimination between sandalwood (San) and non-sandalwood (Nsan) compounds. Indeed, the mean deviation observed is often close to 1 in columns *D*, Q_0 and I_p. There are two main reasons that could explain the important contribution of these descriptors:

(a) The role of *D*, Q_0 and I_p in the establishment of the structure–sandalwood odour relationships, which gives the description of the electronic interactions for the studied molecules with the biological receptor. Recently, Dimoglo

Table 2. Mean deviation Δm (absolute value) between the sandalwood odour observed (0 or 1) and predicted (by the 6 − 8 − 1 configuration network) for each type of compound

Type of compound	*S*	*V*	*O*	*W*	*D*	Q_0	I_p
Acyclic	0·237	0·023	1·000	0·032	1·000	1·000	0·412
Cyclohexyl	0·953	0·476	1·000	0·530	0·985	1·000	0·863
Decalins	0·390	0·776	1·000	0·971	0·782	1·000	0·911
Campholenyl	0·461	0·766	0·847	0·264	1·000	1·000	0·096
Norbornyl	0·476	0·983	0·982	0·262	1·000	1·000	0·714

et al.[48] found the HOMO as a pertinent descriptor in structure–sandalwood odour relationships.

(b) The results achieved with the help of AMPAC showed that the values of D and Q_0 for the compounds containing ketones (Nsan) are higher and lower, respectively, than those of their homologues containing alcohol (generally San). We consider that D and Q_0 play an important role in the discrimination between these two types of compounds. As mentioned above, Q_0 is correlated to log$P$, which accounts for the transport of odorants through the mucus. It is easy to consider Q_0 because it is read directly from the AMPAC output.

As the two main subsets of compounds which have a similar side part (chain or cycle), the descriptors used account for the properties of the bulky part (hydrocarbon) of the molecule which was found as one of the two main structural elements of the santalophore pattern.[8, 13, 14]

CONCLUSION

NNs with a back-propagation algorithm were used for the sandalwood odour classification of chemical compounds. We have shown that NNs can learn the relationship between odour and molecular structures. The predictive power of NN was compared with that of DA. NNs appear able to extract more information from the data than DA. Thus, it is clear that NNs provide a useful method for the analysis of structure–odour relationships and enable us to design new odorant compounds.

The contributions of descriptors to the classification were evaluated. They confirmed the well known role of steric and electronic effects in the establishment of structure–odour relationships.

Although the samples studied included various chemical structures, the established model seemed to be interesting because the predicted odour values are correct for all the subset and a lot of them were in agreement with the observed values. For example, molecules Nos 1, 45 and 65 (which are standard sandalwood odorants, osyrol, sandalore and α-santalol, respectively) were predicted to be strong sandalwoods (their predicted odour values were 1, 1 and 0·91, respectively). Molecule No. 27, representing a non-flexible sandalwood structure, is known to be fairly sandalwood.[8] Its corresponding calculated odour was only 0·54. This showed that the model did not consider such chemical structures as the optimum ones. Molecule No. 31 is less sandalwood than No. 27, because the isopropyl group is smaller than the *tert*-butyl group (found to be the optimum for such structures[8]). The calculated sandalwood odour for molecule No. 31 was 0·52.

All these results showed that the model may be of considerable interest for the design of new sandalwood odorants. The molecular descriptors used were sufficient to describe the main aspects of the molecular structure.

REFERENCES

1. J. E. Amoore, *Ann. NY. Acad. Sci.* **116**, 457–475 (1964).
2. M. G. J. Beets, *Structure–Activity Relationships in Human Chemoreception*. Applied Science, London (1978).
3. G. Ohloff, *Experientia* **42**, 271–279 (1986).
4. M. Chastrette and D. Zakarya, *C.R. Acad. Sci., Ser. II* **307**, 1185–1188 (1988).
5. M. Chastrette, D. Zakarya and A. Elmouaffek, *Eur. J. Med. Chem. Chim. Ther.* **21**, 505–510 (1986).
6. J. N. Narvaez, B. K. Lavine and P. C. Jurs, *Chem. Sens.* **11**, 145 (1986).
7. C. L. Ham and P. C. Jurs, *Chem. Sens.* **10**, 491–505 (1985).
8. M. Chastrette, D. Zakarya and C. Pierre, *Eur. J. Med. Chem.* **25**, 43 (1990).
9. B. P. Eminet and M. Chastrette, *Chem. Sens.* **7**, 293–300 (1983).
10. D. Zakarya, M. Yahiaoui and S. Fkih-Tétouani, *J. Soc. Mar. Chim.* **1**, 14–20 (1993).
11. D. Zakarya, M. Yahiaoui and S. Fkih-Tétouani, *J. Phys. Org. Chem.* **6**, 627–633 (1993).
12. G. Ohloff, in *Fragrance Chemistry. The Science of the sense of smell*, edited by E. T. Theimer, pp. 535–570. Academic Press, New York (1982).
13. E. J. Brunke and E. Klein, in *Fragrance Chemistry. The Science of the Sense of Smell*, edited by E. T. Theimer, pp. 397–429. Academic Press, New York (1982).
14. R. E. Naipawer, K. L. Purzycki, G. W. Schaffer and R. E. Ericksson, in *Essential Oils*, edited by B. P. Mookherjee and C. S. Mussinan. Allured, Wheaton, IL (1981).
15. G. Buchbauer, K. Leonhardsberger, S. Winiwarter and P. Wolschann, *Helv. Chim. Acta* **75**, 174–182 (1992).
16. J.-L. McClelland, D. E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing*, Vol. I. MIT Press, Cambridge, MA (1988).
17. J. A. Burns and G. M. Whitesides, *Chem. Rev.* **93**, 2583–2601 (1993).
18. M. Tusar, J. Zupan and J. Gasteiger, *J. Chim. Phys.* **89**, 1517–1529 (1992).
19. D. Villemin, D. Cherqaoui and J.-M. Cense, *J. Chim. Phys.* **90**, 1505–1519 (1993).
20. D. Villemin, D. Cherqaoui and A. Mesbah, *J. Chem. Inf. Comput. Sci.* **34**, 1288–1293 (1994).
21. D. Dominé, J. Devillers, M. Chastrette and W. Karcher, *SAR QSAR Environ. Res.* **1**, 211–219 (1993).
22. M. Sofan, A. Abdelmegied, M. Pedersen, E. Pedersen and C. Nielsen, *Synthesis* **5**, 516–520 (1994).
23. M. Chastrette, D. Zakarya and J. F. Peyraud, *Eur. J. Med. Chem.* **29**, 343–348 (1994).
24. M. Chastrette and J. Y. De Saint Laumer, *Eur. J. Med. Chem.* **26**, 829–833 (1991).
25. D. Cherqaoui, D. Villemin and V. Kvasnicka, *Chemom. Intell. Lab. Syst.* **24**, 117–121 (1994).
26. N. Bodor, A. Harget and M. J. Huang, *J. Am. Chem. Soc.* **113**, 9480–9483 (1991).
27. J.-M. Cense, B. Diawara, J. J. Legendre and G. Roullet, *Chemom. Intell. Lab. Syst.* **23**, 301–308 (1994).
28. D. Zakarya, L. Farhaoui and S. Fkih-Tétouani, *Tetrahedron Lett.* **35**, 4985–4988 (1994).
29. V. Simon, J. Gasteiger and J. Zupan, *J. Am. Chem. Soc.* **115**, 9148–9159 (1993).
30. J. U. Thomsen and B. Meyer, *J. Magn. Reson.* **84**, 212–217 (1989).

31. E. W. Robb and M. E. Munk, *Mikrochim Acta* **I**, 131–155 (1990).
32. M. E. Munk, M. S. Madison and E. W. Robb, *Mikrochim Acta* **II**, 505–514 (1991).
33. Y. Miyashita, H. Yoshida, O. Yaegashi, T. Kimura, H. Nishiyama and S. Sasaki, *J. Mol. Struct.* (*Theochem*) **311**, 241–245 (1994).
34. B. Curry and D. E. Rumelhart, *Tetrahedron Comput. Methodol.* **3**, 213–237 (1990).
35. L. H. Holley and M. Karplus, *Proc. Natl. Acad. Sci USA* **86**, 152–156 (1989).
36. N. Qian and T. J. Sejnowski, *J. Mol. Biol.* **202**, 865–884 (1988).
37. C. Pierre, *Rapport de Fin d'Études CNAM*. Université Claude Bernard, Lyon (1990).
38. N. Bodor, M.-J. Huang and A. Harget, *J. Mol. Struct.* (*Theochem*) **309**, 259–266 (1994).
39. *HyperChem*, *Release 4, No. 519-10007319.* Hypercube, Waterloo, Ontario.
40. M. J. S. Dewar, E. G. Zoebisc, E. F. Healy and J. J. P. Steward, *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
41. A. Gavezzotti, *J. Am. Chem. Soc.* **105**, 5220–5225 (1983).
42. J. L. McClelland, D. E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing*, Vol. I, pp. 319–362. MIT Press, Cambridge, MA (1988).
43. D. Cherqaoui and D. Villemin, *J. Chem. Soc.*, *Faraday Trans.* **90**, 97–102 (1994).
44. J. A. Freeman and D. M. Skapura. *Neural Networks Algorithms*, *Applications*, *and Programming Techniques*, pp. 89–125. Addison-Wesley, Reading, MA (1991).
45. T. A. Andrea and H. Kalayeh, *J. Med. Chem.* **34**, 2824 (1991).
46. M. Chastrette, D. Crétin and C. El Aïdi, *J. Chem. Inf. Comput. Sci.* **36**, 108–113 (1996).
47. J. Zupan and J. Gasteiger, *Neural Networks for Chemists*. VCH, New York (1993).
48. A. S. Dimoglo, A. A. Beda, N. M. Shvets, M. Yu. Gorbachov, L. A. Kheifits and I. S. Aulchenko, *New J. Chem.* **19**, 149–154 (1995).

