JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 6, 627–633 (1993)

STRUCTURE–ODOUR RELATIONS FOR BITTER ALMOND ODORANTS

DRISS ZAKARYA,[a,b,*] MOHAMED YAHIAOUI[b] AND SOUÄD FKIH-TETOUANI[b]

[a] *Département de Chimie, Faculté des Sciences, Université My Ismaïl, Meknès, Morocco, and*
[b] *Laboratoire de Chimie des Plantes et de Synthèse Organique et Bioorganique, Faculté des Sciences, Université Mohammed V, Rabat, Morocco*

Structure–bitter almond odour relations are established for a set of 65 organic compounds (40 having the bitter almond character) belonging to benzene, pyrrole, thiophene, acyclic and cyclic compounds. Compounds are described using components of autocorrelation method, and the odour is described with a binary variable. Data were analysed using principal component analysis followed by a linear discriminant analysis. The obtained model gives satisfactory classification and prediction of the training set and the test set, respectively. In addition, obtained structure–odour relations were translated to structural elements (and rules) necessary for a stimulus to have bitter almond character.

INTRODUCTION

The conception and synthesis of odorant compounds are very useful to the perfumery and aromatic industries. Design of new active molecules can be based on the structural similarities between them and an existing leader. Nevertheless, this approach is generally very limited by the number of possibilities offered to the chemist. New molecules can also be designed on the basis of knowledge of their specific receptors. In fact, the existing information is not sufficient to allow the establishment of a general theory about olfactory mechanisms. Until more biochemical and biological information is available, it is possible to design new molecules by using structure–odour relations, which also allow an understanding of the mechanisms that govern the studied odour. Thus it was possible to state that ambergris,[1] musk[2,3] and sandalwood[4–6] fragrances were due to the presence in the chemical structure of well defined substructures.

In this paper, we present a simple model to predict correctly the bitter almond fragrance for organic compounds. This was possible by using the components of autocorrelation vectors as molecular descriptors. Starting from this model, we show that it is possible to derive a supplementary structural element that can be added to the well known ones,[7,8] in order to improve understanding of structure–bitteralmond relations and then certain aspects of the olfactory mechanisms.

MATERIAL AND METHODS

Data set and odour coding. We have considered a set of 65 organic compounds (Table 1) derived from different chemical structures (acyclic, benzenic, etc.). Forty of them are bitter almond (BA) odorants.

The studied sample was divided into a training set (40 compounds, 20 BA) and a test set (25 compounds, 20 BA). The bitter almond odour (BAexp.) was described using a dichotomous variable (1 for BA and −1 for non-BA (nBA)). BAcalc. is a discriminant function representing the bitter almond odour, calculated using the obtained model. This function is taken as equal to 1 when its calculated value is positive, and equal to −1 when its calculated value is negative.

Description of the chemical structure and data analysis. Each molecule was described using components of the modified autocorrelation vectors[13] based on the Van der Waals volume (V) and surface (S).[13,14] These vectors account for the size and the shape of the molecule, and electronegativity (E)[15] (using the classical autocorrelation method[16]) in order to take into account the electronic aspect. The general relation used to calculate the autocorrelation components is:

$$C_k = \sum_{\substack{k \\ j \geqslant i}} [f(i) \cdot f(j)]^x \qquad (1)$$

x is equal to 0·5 in the case of the modified autocorrelation method and equal to 1 in the case of the classical

* Author for correspondence.

0894–3230/93/110627–07$08.50

Received 10 April 1993
Accepted 5 July 1993

Table 1. Chemical structures of the studied compounds and their odours

No.	Compound	BAexp.	Ref.	BAcalc.
Training set				
1	Benzaldehyde	1	7	1 (0·92)
2	2-Methylbenzaldehyde	1	7	1 (0·94)
3	3-Methylbenzaldehyde	1	7	1 (0·40)
4	2-Ethylbenzaldehyde	1	7	1 (0·76)
5	3-Ethylbenzaldehyde	1	7	−1 (−0·30)
6	2-Chlorobenzaldehyde	1	7	1 (0·86)
7	2-Fluorobenzaldehyde	1	8	1 (0·90)
8	2,4-Dimethylbenzaldehyde	1	8	−1 (−0·42)
9	4-Methylbenzaldehyde	1	9	1 (0·03)
10	Furfural	1	7	1 (0·85)
11	5-Methylfurfural	1	7	1 (0·50)
12	Thiophene-2-carboxaldehyde	1	7	1 (1·32)
13	2-Nitropyrrole	1	8	1 (0·80)
14	Pyridine-3-carboxaldehyde	1	9	1 (0·60)
15	3-Chlorobenzaldehyde	1	7	1 (0·28)
16	2-Methylbut-2-en-1-al	1	7	1 (0·11)
17	2-Ethylbut-2-en-1-al	1	7	1 (0·93)
18	2-Ethylpent-2-en-1-al	1	7	1 (0·95)
19	2-Methyl-3-cyclopropylprop-2-en-1-al	1	10	1 (1·17)
20	1-Cyclooct-1-ene carboxaldehyde	1	7	1 (1·22)
21	4-Isopropylbenzaldehyde	−1	7	−1 (−0·71)
22	4-Nitrobenzaldehyde	−1	11	−1 (−1·18)
23	4-Methoxybenzaldehyde	−1	12	−1 (−0·50)
24	4-Chlorobenzaldehyde	−1	11	−1 (−0·02)
25	3-Phenylprop-2-en-1-al	−1	7	−1 (−1·26)
26	4-Methylbenzonitrile	−1	11	−1 (−0·06)
27	4-Bromobenzonitrile	−1	11	−1 (−0·13)
28	3-Bromonitrobenzene	−1	11	−1 (−0·51)
29	4-Nitrobenzonitrile	−1	11	−1 (−1·28)
30	4-Ethoxybenzaldehyde	−1	9	−1 (−0·69)
31	4-Ethoxyacetophenone	−1	9	−1 (−1·25)
32	4-Methylacetophenone	−1	9	−1 (−1·34)
33	4-Methoxy-3-methylbenzaldehyde	−1	9	−1 (−1·27)
34	2-Methoxybenzaldehyde	−1	9	1 (0·33)
35	2-Methylprop-2-en-1-al	−1	12	−1 (−0·76)
36	But-2-en-1-al	−1	7	−1 (−0·47)
37	Hexa-2,4-dien-1-al	−1	7	−1 (−0·66)
38	Hexa-2-en-1-al	−1	7	−1 (−0·21)
39	4-Isopropenyl-1-carboxaldehyde 1-Cyclohexene	−1	12	−1 (−0·61)
40	2,6,6-Trimethyl-1-carboxaldehyde Cyclohexa-1,3-dien	−1	12	−1 (−0·22)
Test set				
41	Nitrobenzene	1	7	1
42	Benzonitrile	1	7	1
43	2-Methylnitrobenzene	1	8	1
44	3-Methylnitrobenzene	1	8	−1
45	Acetophenone	1	7	1
46	2-Methylbenzonitrile	1	8	1
47	3-Methylbenzonitrile	1	8	1
48	2-Fluorobenzonitrile	1	8	1
49	3-Methylfurfural	1	7	1
50	4-Methylfurfural	1	7	1
51	2-Nitrofurane	1	7	1
52	2-Cyanofurane	1	8	1
53	2-Nitrothiophene	1	8	1

continued

Table 1. (*Continued*)

No.	Compound	BAexp.	Ref.	BAcalc.
54	3-Nitrothiophene	1	8	1
55	N-Methyl-2-nitropyrrole	1	8	1
56	2-Methylpent-2-en-1-al	1	7	1
57	Cyclohex-1-en-1-carboxaldehyde	1	7	1
58	Cyclopent-1-en-1-carboxaldehyde	1	7	1
59	Cyclohept-1-en-1-carboxaldehyde	1	7	1
60	Naphtalene-1-carboxaldehyde	1	8	1
61	4-Chlorobenzonitrile	−1	11	−1
62	4-Methylnitrobenzene	−1	11	−1
63	4-Chloronitrobenzene	−1	11	−1
64	4-Bromonitrobenzene	−1	11	−1
65	4-Methoxynitrobenzene	−1	11	−1

method. C_k is the autocorrelation component corresponding to the topological distance k (smallest number of bonds) between atoms i and j. f(i) and f(j) are the contributions of atoms i and j to the considered property (volume, electronegativity, etc).

Eight components were calculated ($k = 0$ to 7) for each autocorrelation vector (V, S and E). Thus, each molecule was described by 24 (8×3) molecular descriptors.

Statistical analyses were conducted by means of principal component analysis (PCA),[17] followed by a discriminant analysis (DA).[18] PCA gave a general view of the data set while DA allowed selection of the most significant variables for discriminating between BA and nBA compounds. PCA was conducted using STATITCF software,[19] and DA using a simple program.

RESULTS AND DISCUSSIONS

Principal component analysis

The training sample, initially represented by a set of points in a 24-dimensional space, can be represented in two-dimensional space F1–F2. In the present study, F1 and F2 components account for 49·8% and 23·2%, respectively, representing a correct percentage (73%) of the total variance.

The projection of the cloud of points on the plane F1-F2 shows that the separation of the classes of active and inactive compounds is satisfying. We noted that the points representing standard BA compounds, such as benzaldehyde and furfural, are situated in a region on the right of the plane without points representing nBA compounds. (Figure 1).

To test the validity of the PCA classification, we plotted the points representing the tested molecules on F1–F2 plane, after calculation of their coordinates using the prediction procedure of the PCA program. We noted that BA and nBA compounds of the test set were correctly classified into their corresponding categories.

PCA classification is interesting, and can be used as a model for the prediction of bitter almond fragrance for unknown structures. However, it is probable that only few variables, among the 24 used, are responsible for the obtained classification. This suggests the use of a method such as DA, which allows the classification of active and inactive compounds using the statistically significant variables only.

Discriminant analysis

The statistical discrimination between active and inactive compounds of the training set was performed by means of a linear discriminant analysis leading to the following model.

$$\text{Bacalc.} = 0{\cdot}112V_1 + 0{\cdot}607S_3 - 0{\cdot}095V_3 - 0{\cdot}578S_6 - 0{\cdot}047E_5 - 4{\cdot}31 \quad (2)$$

V_1 and V_3 are components of the autocorrelation vector V corresponding to $k = 1$ and $k = 3$; S_3 and S_6 are components of the autocorrelation vector S corresponding to $k = 3$ and $k = 6$; E_5 is a component of the autocorrelation vector E corresponding to $k = 5$.

All the coefficients associated with the variables were statistically significant ($p < 0{\cdot}03$). Table 2 shows the percentages of good classification and good prediction using the above model. These results show that the obtained model is good for predicting the bitter almond

Table 2. Percentages of good classification (training set) and good prediction (test set)

	Bitter almond	Non-bitter almond
Training set	90	95
Test set	95	100

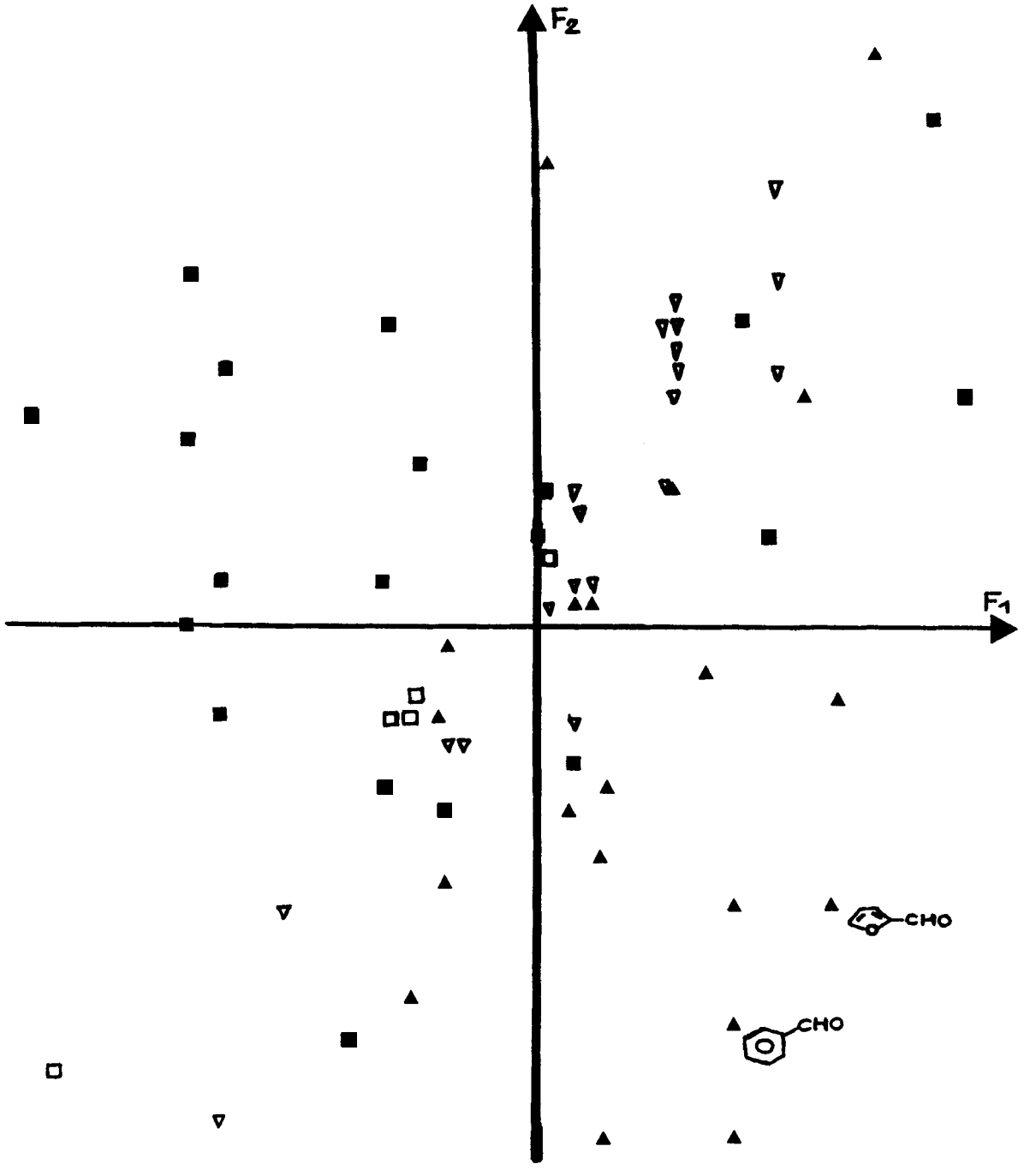

Figure 1. Projection of the cloud of points (compounds) on the F1–F2 plane. ▲, BA (training set); ▽, BA (test set); ■, nBA (training set); ◻, nBA (test set)

character for unknown molecules having a chemical structure similar to the studied ones.

In the next step, we attempt translation of the molecular information provided by the model to structural information easily understood by the chemist.

Despite the general difficulty in obtaining a satisfactory interpretation of the autocorrelation components, it has been shown recently[20] that the components with a small index are essentially correlated to the size of the molecule while those with a higher index account for the molecular branching. Obviously, when the studied molecules have similar chemical structures, the components of higher index also take into account the size of the molecule. Here, the obtained model comprises V_1, which is well correlated with the size of the molecule (which is represented by the number of atoms, N_{atoms}):

$$V_1 = 7{\cdot}632N_{atoms} - 3{\cdot}271 \qquad (3)$$
$$n = 40 \quad r = 0{\cdot}939 \quad s = 4{\cdot}74$$

Considering its associated coefficient ($K_1 = 0{\cdot}112$), the increase of V_1, then the size of the molecule, favours the presence of bitter almond character; however, V_1 has a minor contribution (12%, calculated using Gore's method[21]).

For better simplification, the components S_3 and V_3 are considered in the term $K_2S_3 - K_3V_3$ ($K_2 = 0{\cdot}607$ and $K_3 = -0{\cdot}095$) because they account for the same aspect of the molecule. Their contributions were 52% and 22·8%, respectively, with a total of 74·8%. As the studied molecules have similar chemical structures, the $K_2S_3 - K_3V_3$ term is also correlated with the size, as shown by equation (4).

$$K_2S_3 - K_3V_3 = -0{\cdot}439N_{atoms} + 2{\cdot}563 \qquad (4)$$
$$n = 40 \quad r = 0{\cdot}882 \quad s = 0{\cdot}397$$

Finally, the $K_2S_3 - K_3V_3$ term accounts for the size and the shape of the molecule, and considering its high contribution (74·8%), it seems sufficient alone to explain the origin of the bitter almond fragrance.

Components E_5 and S_6 have minor contributions (5·6% and 6·7%, respectively); their decrease slightly favours the bitter almond odour. This means that chemical structures, including a lot of atoms pair compared to the leader one (benzaldehyde) with a topological distance of about five to six bonds, have a slightly diminished bitter almond odour. This is probably because compounds 5 and 8 were calculated by the model as nBA odorants and compound 9 had a discriminant function (BAcalc.) almost equal to zero.

Considering the E_5 contribution, it seems surprising that the electronic aspect of the molecule is not taken into account sufficiently by the obtained model. This is because all the studied compounds have relatively similar chemical structures. In addition, some components of the electronegativity vector were well correlated with components of the Van der Waals volume or surface. This is due to the chemical constitution of the studied set. Thus, we have obtained a model closely similar to equation (2) including only electronegativity components. This model was less efficient than equation (2) (95% and 75% of good classification of BA and nBA, respectively). The model essentially includes components related to the volume and surface properties of the molecule (with a total contribution of 94·4%), expressing the size and the shape. Thus, compound 34 was incorrectly classified as a BA compound. Its chemical structure can he derived from those of compound 4 (BA) by substituting a $-CH_2$ group by an ether function (–O–).

It is well known that the bitter almond fragrance is due to the presence of a hydrogen-bonding acceptor (HBA) functional group (CHO, NO_2, CN, ...) in conjugation with a double bond[7,22] (pattern 1, Figure 2). All the studied compounds possess pattern 1 (or equivalent), then the obtained model is a function between the structural environment of this pattern and the bitter almond odour. Recently it has been shown[22] that this environment can be considered as the third structural element necessary for the presence of bitter almond fragrance, as shown in Figure 3. Figure 3 shows that the optimum size of the environment corresponds to that of the benzaldehyde molecule. Indeed, this molecule is considered as the strongest bitter almond odorant. Note that a size outside certain limits destroys the odour, despite the presence of pattern 1.

As has been mentioned, our model expresses essentially the size and the shape of the studied molecules. Since all these include pattern 1 (or equivalent), then the model accounts for the influence of the size and the shape of the environment on the bitter almond odour, as shown by Figure 4. Thus, we note that molecules with a small number of atoms (size) have a negative discriminant function, as well as those having a high number of atoms. Benzaldehyde (strong bitter almond) is located in the optimum region of the cloud of points. This is clearly shown when we consider a set of similar chemical structures (Figure 5).

According to the cited observations, it seems that the presence of the bitter almond fragrance requires a substructure such as pattern 1 in conjunction with an environment having a size between a minimum and a maximum value that can be determined from the molecules cited in Figures 4 and 5. The optimum size is close to that of the benzaldehyde. All the obtained

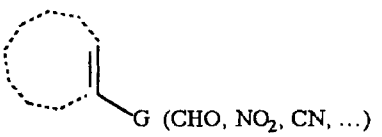

Figure 2. Structure of pattern 1

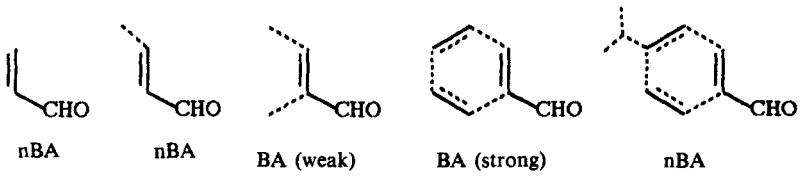

Figure 3. Dependence of bitter almond fragrance on the size of the environment of pattern 1

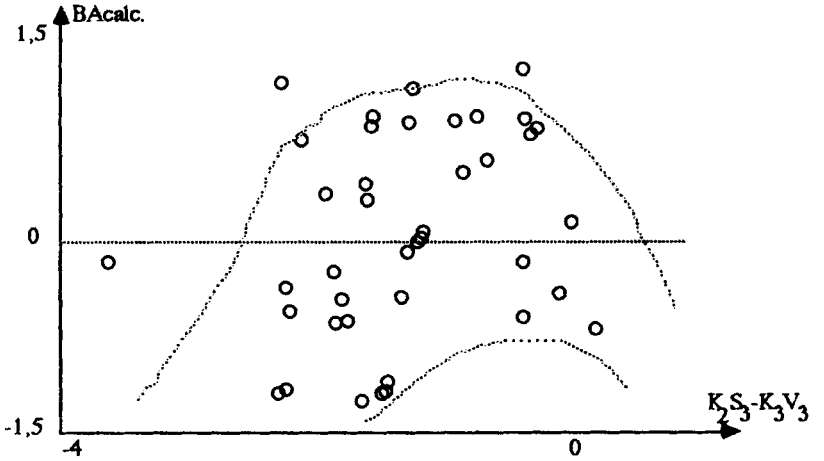

Figure 4. Variation of BAcalc. with $K_2S_3 - K_3V_3$ (for all compounds of the training set)

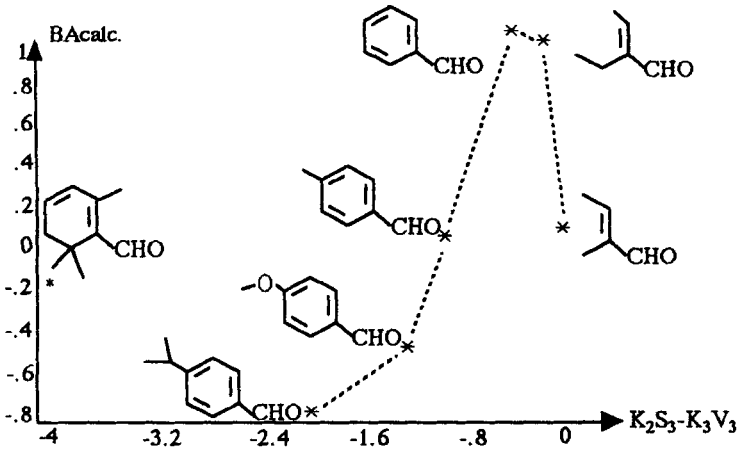

Figure 5. Variation of BAcalc. with $K_2S_3 - K_3V_3$ for a set of molecules

Weak bitter almond Strong bitter almond

Figure 6. Dependence of bitter almond intensity on the number of double bonds

results suggest the establishment of more efficient rules for the prediction of the bitter almond odour for organic compounds similar to those studied. These rules are cited below.

1. The chemical structure must include an HBA group such as CHO, NO_2, CN or equivalent.
2. There must be a double bond, conjugated with the HBA group. Intensity of the bitter almond fragrance seems to be increased by the introduction of supplementary double bonds[7] (Figure 6).
3. There must be a substructure linked to the double bond, having a size close to that of benzaldehyde, with minimum and maximum limits.

The conditions required for the development of bitter almond fragrance can be translated into an interaction model. However, this is possible only by investigation of supplementary structural information (molecular orbitals, atomic charges, conformational study, ...).

In conclusion, we have shown in this work that it is possible to establish a significant relation between bitter almond fragrance and chemical structure of organic compounds. This result is in good agreement with previous studies[23,24] showing that bitter almond seems to be a well defined odour for the human olfactory system.

Structure–odour relations were translated to structural information about the structural elements (rules) that a given stimulus is required to develop bitter almond fragrance. For more information, a molecular mechanics study of the cited compounds is now in progress.

ACKNOWLEDGEMENTS

We thank Dr A. Mhamdi (from IAV-Rabat) for assistance concerning statistical analyses. We also thank the referees for their interesting suggestions.

REFERENCES

1. G. Ohloff, *Experientia* **42**, 271 (1986).
2. M. G. J. Beets, *Structure–Activity Relationships in Human Chemoreception*, Applied Science Publishers, London (1978).
3. M. Chastrette and D. Zakarya, *C.R. Acad. Sci. Paris* **307**(II), 1185 (1988).
4. E. J. Brunke and E. Klein, Chemistry of Sandlewood Fragrance, in *Fragrance Chemistry*, edited by E. T. Theimer, Academic Press, New York (1982).
5. M. Chastrette, D. Zakarya and C. Pierre, *Eur. J. Med. Chem.* **24**, 433 (1990).
6. G. Buchbauer, K. Leonhardsberger, S. Winiwarter and P. Wolschann, *Helv. Chim. Acta* **75**, 174 (1992).
7. H. Boelens, *Cosmetic and Perfumery* **89**, 70 (1974).
8. A. Eriksson, P. Lindner and O. Martensson, *Molecular Properties and Odour. A study of the effect of substituents in some five and six membered aromatic rings*, Uppssala University (1977).
9. S. Arctander, *Perfume and Flavor Chemicals*, Monclair NJ, (1969).
10. H. Boelens and J. Heydel, *Chem. Zeit.* **97**, 1 (1973).
11. A. Eriksson, P. Lindner and O. Martensson, *J. Theor. Biol.* **90**, 477 (1981).
12. E. Guenther, *The Essential Oils*, Vol. II, Van Nostrand Reinhold, New Jersey (1945).
13. M. Chastrette, F. Tiyal, J. F. Peyraud and D. Zakarya, *C.R. Acad. Sci. Paris* **314**(II), 461 (1992).
14. A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
15. L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca (1960).
16. P. Broto, G. Moreau and C. Vandycke, *Eur. J. Med. Chem. -Chim. Ther.* **19**, 61 (1984).
17. L. Lebart, A. Morineau and J. P. Fenelon, *Traitement des Données Statistiques*, Dunod, Paris (1979).
18. P. Dagnélie, *Analyses Statistiques à Plusieurs Variables*, Presse Agron, Gembloux, Belgium (1979).
19. STATITCF software, Institut Technique des Céréales et des Fourrages, Paris, France (1987).
20. D. Zakarya, *New J. Chem.* **16**, 1039 (1992).
21. W. L. Gore, *Statistical Methods for Chemical Experimentation*, Interscience, New York, (1952).
22. D. Zakarya, M. Yahiaoui and S. Fkih-Tetouani, *J. Soc. Mar. Chim.* **1**, 14 (1992).
23. M. Chastrette, A. Elmouaffek and D. Zakarya, *C.R. Acad. Sci. Paris, série II*, **303**, 1209 (1986).
24. M. Chastrette, A. Elmouaffek and P. Sauvegrain, *Chem. Senses* **13**, 295 (1988).