Eur J Med Chem (1995) 30, 679–686

Tetralin, indan and nitrobenzene compound structure–musk odor relationship using neural networks

M Chastrette, C El Aïdi, JF Peyraud

Laboratoire de chimie organique physique et synthétique, URA 463, Université Claude-Bernard-Lyon 1, 69622 Villeurbanne Cedex, France

(Received 26 January 1995; accepted 3 May 1995)

Summary — Structure–musk odor relationships were established by means of a neural network (NN) on a sample of 105 molecules comprising nitrobenzene compounds, carbonyl tetralins and carbonyl indans. Each molecule was described by eight variables (six steric hindrance descriptors and two electronegativity descriptors, coding six substituents on a common underlying skeleton (benzene ring with a tertiobutyl or pseudo-tertiobutyl group in the *meta* position relative to a nitro or carbonyl group)). Odor was coded by a binary variable. Seventeen molecules could be coded by two different descriptions, and so the back-propagation NN was trained on 122 descriptions. The NN gave a better classification (98.1%) than that obtained by discriminant analysis (65.7%). To test the reliability of the NN approach, training was made on 11 training subsets of 111 or 110 descriptions with a correct classification rate of 96.4 to 98.6%. For the 11 corresponding test subsets of 11 or 12 descriptions a correct prediction rate of 80% was obtained for the 105 molecules. The contribution of each descriptor was evaluated. The results show the possibility of establishing structure–musk odor relationships for molecules with different structures and confirm the previously mentioned effect of steric hindrance in the *ortho* position of the functional groups interacting with the receptor by hydrogen bonding.

structure–activity relationship / musk odor / tetralin / indan / nitrobenzene compound / neural network / back-propagation / descriptor contribution

Introduction

The perception of an odor is the outcome of a complex process that begins with the action of an odorous compound on receptors in the olfactory mucosa. Although, *a priori*, its appears in different chemical families, the characteristic odor of musk might well correspond [1] to a set of relatively homogeneous receptors. Several different musk structure–odor relationships have been established for all or part of a set of more than 360 compounds.

Using a sample of 148 bicyclic molecules, Jurs *et al* [2] could associate the odor of musk with the simultaneous presence of a carbonyl group and two quaternary carbons in the molecule at fixed distances from one another. More recently, Bersuker *et al* [3] studied a sample of 362 musks using electron-topological matrices of contiguity (ETMC) and identified two fragments responsible for musk fragrance. Chastrette and Zakarya have, for their part, demonstrated the compatibility of the three patterns responsible for musk odor in three major musk families by taking into account the geometic contraints of hydrogen bonding within the framework of the HBD (hydrogen bonding and dispersion) theory [4].

Several applications in chemistry [5–8] and, in particular, new approaches to the structure–odor relationship [1, 9–11] have been developed using neural networks (NN) [12–16]. Chastrette and Saint-Laumer [17] modelled the structure–odor relationship of 79 nitrobenzene compounds using a multilayer back-propagation network [18, 19]. Each molecule was described as a skeleton with five substituents, each of which was coded with six variables, giving a total of 15 Van-der-Waals volume descriptors and 15 electronegativity descriptors. It was possible to obtain a classification rate of 94% and a prediction rate of 77% with the network.

Chastrette and Zakarya [20] also used the same method, with a three-layer NN, to study a series of 53 tetralins and 15 carbonyl indans. Each molecule was coded with seven variables (six steric hindrance descriptors and one electronegativity descriptor) describing substituents of the tetralin or indan skeleton. The network in its training phase on the 53 tetralins correctly classified all of them (100%). All the indans were subsequently correctly predicted in the test phase and it was possible to measure the relative contribution of each descriptor to the classification.

The aim of this study was to take the previous results (obtained with families of musks taken separately) one step further by extending them to a set of musks comprising carbonyl tetralins, carbonyl indans, and nitrobenzene compounds. Such an approach is quite conceivable considering the analogy that can be made between the hydrogen bond acceptor capacities of RCO and NO_2 groups within the framework of the HBD theory [4]. All the compounds in the set had a tertiobutyl group or its equivalent in the *meta* position of a nitro group for the nitrobenzene compounds or a carbonyl group for the indans and tetralins. In order to be able to compare the results obtained with a linear method with those obtained with a non-linear approach based on NN, we also carried out discriminant analysis [21, 22] on the sample used. Thereafter, we sought to measure the contribution of each descriptor to the classification.

Materials and methods

Materials

The sample comprised 105 compounds (table I): 60 carbonyl tetralins 32 of which were musks; 20 carbonyl indans 17 of which were musks; and 25 nitrobenzene compounds with a tertiobutyl group in the *meta* position of the NO_2, 14 of which were musks. The tetralins and indans were taken from the sample of Zakarya *et al* [20] and the nitrobenzene compounds from that of Chastrette and Saint-Laumer [17]. The olfactory properties of these compounds have been described previously [17, 20]. The odor was coded using one criterion, absence or presence, without indication of intensity.

Description of the structure

The molecules were described from a common underlying skeleton comprising a benzene cycle with a tertiobutyl or pseudo-tertiobutyl group (R_5, fig 1a) and a nitro or carbonyl group in the *meta* position (fig 1a). The choice of samples was largely a function of the data available in the literature, the compatibility of molecules with the underlying skeleton, and the compatibility of substituents with the system of description adopted.

The benzene nucleus can have six substituents (fig 1a), including the RXO group in position 1 (X = C or N; R and O are denoted by R_1 and O_1 respectively). In the first instance, the substituent R_5 was not taken into account in the description of the skeleton which comprised eight descriptors: six steric hindrance descriptors (D_1 to D_4, D_6 and Do_1 (corresponding to the respective hindrances of groups R_1 to R_4, R_6 and to oxygen O1)), and two electronegativity descriptors (D_7 and D_8 (corresponding to the electronegativities of groups R_2 and R_1)).

In order to be able to obtain compatible descriptions of substitution for benzene compounds and for indans or tetralins, we introduced a fictitious cut for the latter as shown in figures 1b and 1c.

The structure of the nitrobenzene compounds was described as shown in figure 1d. However, for the 17 compounds with a second RXO group in *meta* position of the tertiobutyl substi-

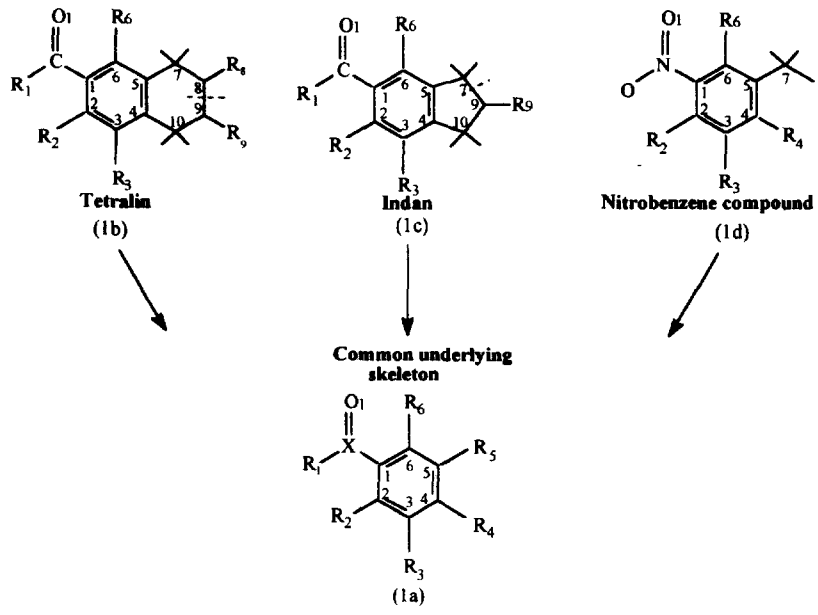

Fig 1. Underlying skeleton and substituents for the three families of musks. X = C or N. R_1, R_2, R_3, ... R_6 = O, alkyl, alkoxy, halogen *etc.*

tuent, two different descriptions were used. It was sufficient for one of the descriptions to lead to a correct classification for the molecule to be considered as being correctly classified (or predicted). In effect, the two different descriptions of the same molecule relate to two different orientations relative to the receptor. If only one of these two orientations corresponds to a muskophore pattern, the receptor will recognize it.

The substituent hindrance was described with effective upsilons U_{ef} or, where impracticable, with the minimum values U_{min} of Charton's upsilon [23]. The alkyl groups not studied by Charton were assimilated to similar groups (table II). Thus, for molecule 23, substituent $-C(Me)_2C(Me)_2-$ obtained after the fictitious cutting shown in figure 1b, was assimilated to $-C(Me)_2tBu$ (U_{ef} = 2.43). For tetralin and indan molecules, the substituent located at position 4 (fig 1b and 1c), $-C(Me)_2CH_2-$, was assimilated to a tertiobutyl (U_{ef} = 1.24) and substituents $-C(Me)_2CH_2(Me)-$, $-C(Me)_2CHMecy-$, $-C(Me)_2CHEt$ (table I) to a tertioamyl (U_{ef} = 1.63).

The Mecy substituent representing a $-CH_2$ cyclic group was considered as a methyl. For the NO_2 group, the two oxygens were considered as sp^3 oxygen for R_1 and sp^2 oxygen for O_1. The U values of the NO_2 oxygen atoms (R_1 and O_1; U_{est} = 0.4) and the RCO oxygen atom (O_1; U_{est} = 0.45) were estimated by analogy with the U_{ef} values of the NO_2 and RCO groups. Finally, in order to take into account the steric effects of the methyl in the methoxy group in the *ortho* position of bulky substituents, the U_{ef} value of the methoxy substituent was increased from 0.36 to 0.42.

The electronic effects of substituents R_1 and R_2 were described with Boyd and Boyd group electronegativity values [24] and, where impracticable as for halogens and NO_2 oxygen R_1, with Pauling electronegativity [25]. Following the methodology of a previous work [20], the electronegativity of the substituents was adjusted by subtracting the hydrogen electronegativity value. The NO_2 oxygen atom R_1 was considered as an sp^2 atom and was finally attributed the electronegativity value of 1.53.

Neural network

We used a three-layer back-propagation learning network to study the sample described above. The connection between the three different layers was complete (fig 2).

The number of neurons in the input layer was equal to the number of molecular descriptors whereas the output layer had only one neuron. The number of neurons in the hidden layer was determined by trial and error. The activation sigmoid function was of the form: $f(x) = a/(1 + \exp(-b(x - c))$ where a is the upper limit of the output values (0, a), b the non-linearity parameter and c the threshold value.

The olfactory activity represented by the output of the network was coded 0.8 if the molecule was musky and 0.2 if the molecule was not musky. A musky compound was considered as correctly classified (or predicted) if the value of the output neuron corresponding to its activity was higher than a threshold value (c = 0.5).

Work with the network was divided into two phases: a learning (classification) phase and a prediction (test) phase. The learning consisted in reducing the error resulting from the comparison between the output computed by the network and the desired output, by adjusting the weights inside the network through an iterative process of gradient back-propagation (fig 2). The error was of the form:

$$E = \sum_k (Sd_k - Sc_k)^2$$

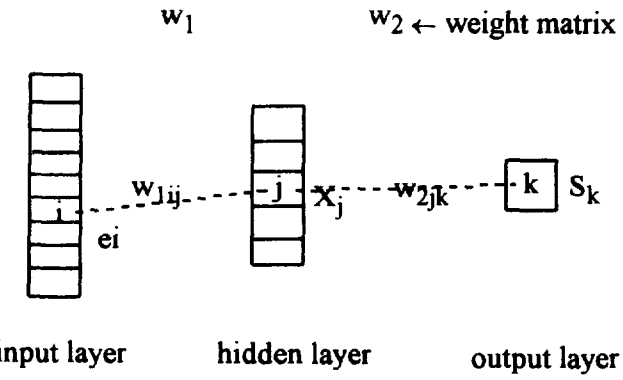

Fig 2. Structure of the multilayer network. $X_j = f(a_j)$; $a_j = \Sigma_i w_{1ij}\, e_i$; $S_k = f(b_k)$; $b_k = \Sigma_j w_{2jk}\, X_j$; w_{1ij}, w_{2jk}: elements of $\mathbf{W}_1$ and $\mathbf{W}_2$ weight matrices; f = activation function.

Sd_k = desired output, Sc_k = computed output.

All 105 molecules in the sample went through the classification phase, with 88 being considered in one description and 17 in two, which amounted to considering 122 descriptions (sample A). The network was trained on the 122 descriptions. The reported results were obtained after 5000 cycles.

In the prediction phase there were 11 BL (B, Learning) reduced samples (ten with 111 descriptions and one with 110) and 11 BT (B, Test) corresponding test samples (ten with 11 descriptions and one with 12). The descriptions of the BT samples were those withdrawn from sample A for the purpose of constituting the BL samples.

The network was successively trained on each of the 11 BL reduced samples, then used in the best configuration to predict the 11 test samples. Each description of sample A was therefore predicted once.

Results and discussion

Classification

The best configuration obtained contained eight neurons on each of the input and hidden layers and only one on the output layer (8-8-1), with the following activation function parameter values: a = 1, b = 0.3 and c = 0.5.

Sample A

The classification yielded 98.1% (103/105) of correctly classified molecules. The incorrectly classified descriptions were Nos 60, 111, 112 and 114 for the musky compounds and descriptions 67, 69 and 77 for the non-musky compounds. However, molecules 59, 62, 67 and 77 corresponding to descriptions 111, 114, 116 and 121 should be considered as correctly classi-

Table I. Substituents and the coding of the 122 descriptions (sample A) corresponding to the 105 molecules studied.

Description[a]	*Family*[b]	*Molecule*	R_1[c]	*Position 2*	*Position 6*	*Position 3*[d]	O_1[e]	*Position 4*[f]	*Odor*[g]
1	T	1	H	H	H	H	RCO	$C(Me)_2CH_2$	0.2
2	T	2	H	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
3	T	3	H	Et	H	H	RCO	$C(Me)_2CH_2$	0.8
4	T	4	H	*i*Pr	H	H	RCO	$C(Me)_2CH_2$	0.8
5	T	5	H	OMe	H	H	RCO	$C(Me)_2CH_2$	0.2
6	T	6	H	H	H	Me	RCO	$C(Me)_2CH_2$	0.2
7	T	7	H	H	H	OMe	RCO	$C(Me)_2CH_2$	0.2
8	T	8	H	Me	Me	H	RCO	$C(Me)_2CH_2$	0.8
9	T	9	H	H	Me	Me	RCO	$C(Me)_2CH_2$	0.2
10	T	10	H	Me	H	Me	RCO	$C(Me)_2CH_2$	0.8
11	T	11	H	Me	OMe	H	RCO	$C(Me)_2CH_2$	0.8
12	T	12	H	Et	OMe	H	RCO	$C(Me)_2CH_2$	0.8
13	T	13	H	H	OMe	Me	RCO	$C(Me)_2CH_2$	0.8
14	T	14	H	Me	H	Mecy	RCO	$C(Me)_2CH_2$	0.8
15	T	15	H	Me	Me	Me	RCO	$C(Me)_2CH_2$	0.2
16	T	16	H	Me	OMe	Me	RCO	$C(Me)_2CH_2$	0.8
17	T	17	H	H	H	H	RCO	$C(Me)_2CHMe$	0.8
18	T	18	H	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
19	T	19	H	Me	Me	H	RCO	$C(Me)_2CHMe$	0.8
20	T	20	H	Me	Me	H	RCO	$C(Me)_2CH_2$	0.8
21	T	21	H	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
22	T	22	H	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
23	T	23	H	Me	H	H	RCO	$C(Me)_2C(Me)_2$	0.2
24	T	24	H	Me	H	H	RCO	$C(Me)_2CHMecy$	0.8
25	T	25	Me	H	H	H	RCO	$C(Me)_2CH_2$	0.2
26	T	26	Me	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
27	T	27	Me	Et	H	H	RCO	$C(Me)_2CH_2$	0.8
28	T	28	Me	*i*Pr	H	H	RCO	$C(Me)_2CH_2$	0.2
29	T	29	Me	OMe	H	H	RCO	$C(Me)_2CH_2$	0.2
30	T	30	Me	F	H	H	RCO	$C(Me)_2CH_2$	0.2
31	T	31	Me	Cl	H	H	RCO	$C(Me)_2CH_2$	0.2
32	T	32	Me	Br	H	H	RCO	$C(Me)_2CH_2$	0.2
33	T	33	Me	CN	H	H	RCO	$C(Me)_2CH_2$	0.2
34	T	34	Me	CO_2Me	H	H	RCO	$C(Me)_2CH_2$	0.2
35	T	35	Me	CH_2OMe	H	H	RCO	$C(Me)_2CH_2$	0.2
36	T	36	Me	OH	H	H	RCO	$C(Me)_2CH_2$	0.2
37	T	37	Me	H	H	Me	RCO	$C(Me)_2CH_2$	0.2
38	T	38	Me	Me	OMe	H	RCO	$C(Me)_2CH_2$	0.2
39	T	39	Me	Me	Me	H	RCO	$C(Me)_2CH_2$	0.2
40	T	40	Et	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
41	T	41	Et	Et	H	H	RCO	$C(Me)_2CH_2$	0.8
42	T	42	Et	*i*Pr	H	H	RCO	$C(Me)_2CH_2$	0.2
43	T	43	*i*Pr	Et	H	H	RCO	$C(Me)_2CH_2$	0.2
44	T	44	Me	H	H	H	RCO	$C(Me)_2CHMe$	0.8
45	T	45	Me	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
46	T	46	Me	Me	H	H	RCO	$C(Me)_2CHEt$	0.8
47	T	47	Me	Me	Me	H	RCO	$C(Me)_2CHMe$	0.2
48	T	48	Me	Me	H	Mecy	RCO	$C(Me)_2CH_2$	0.8
49	T	49	Me	*i*Pr	H	H	RCO	$C(Me)_2CHMe$	0.2
50	T	50	Me	Me	H	H	RCO	$C(Me)_2CHMecy$	0.8
51	T	51	Me	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
52	T	52	Me	Me	Me	H	RCO	$C(Me)_2CH_2$	0.2
53	T	53	Me	Et	H	H	RCO	$C(Me)_2CHMe$	0.8
54	NB	54	O	H	OMe	NO_2	NO_2	H	0.8
55	NB	55	O	Me	OEt	NO_2	NO_2	H	0.8
56	NB	56	O	Et	OMe	NO_2	NO_2	H	0.8
57	NB	57	O	Et	OEt	NO_2	NO_2	H	0.8
58	NB	58	O	*i*Pr	OMe	NO_2	NO_2	H	0.8

Table I. *(Continued.)*

Description[a]	*Family*[b]	*Molecule*	R_1[c]	*Position 2*	*Position 6*	*Position 3*[d]	O_1[e]	*Position 4*[f]	*Odor*[g]
59	NB	59	O	OMe	Br	NO_2	NO_2	H	0.8
60	NB	60	O	OMe	OMe	NO_2	NO_2	H	0.8
61	NB	61	O	H	OMe	CHO	NO_2	H	0.8
62	NB	62	O	H	OMe	*t*Bu	NO_2	H	0.8
63	NB	63	O	H	OMe	H	NO_2	H	0.8
64	NB	64	O	Et	H	NO_2	NO_2	H	0.8
65	NB	65	O	OMe	Br	H	NO_2	Br	0.8
66	NB	66	O	OMe	Me	NO_2	NO_2	H	0.2
67	NB	67	O	H	OMe	COMe	NO_2	H	0.2
68	NB	68	O	Me	OMe	*t*Bu	NO_2	H	0.2
69	NB	69	O	Me	OMe	Me	NO_2	H	0.2
70	NB	70	O	Me	H	NO_2	NO_2	H	0.2
71	NB	71	O	OMe	NO_2	Me	NO_2	NO_2	0.2
72	NB	72	O	Me	*i*Pr	Me	NO_2	NO_2	0.2
73	NB	73	O	Me	OMe	NO_2	NO_2	H	0.8
74	NB	74	O	H	OEt	NO_2	NO_2	H	0.8
75	NB	75	O	OMe	H	NO_2	NO_2	H	0.2
76	NB	76	O	*t*Bu	OMe	NO_2	NO_2	H	0.2
77	NB	77	O	Br	OMe	NO_2	NO_2	H	0.2
78	NB	78	O	OMe	H	CHO	NO_2	H	0.2
79	I	79	O	Me	H	NO_2	NO_2	$C(Me)_2CH_2$	0.8
80	I	80	O	Me	H	NO_2	NO_2	$C(Me)_2CHMe$	0.8
81	I	81	O	Et	H	NO_2	NO_2	$C(Me)_2CH_2$	0.8
82	I	82	O	Me	H	NO_2	NO_2	$C(Me)_2CHEt$	0.8
83	I	83	O	Et	H	NO_2	NO_2	$C(Me)_2CHMe$	0.8
84	T	84	O	Me	H	NO_2	NO_2	$C(Me)_2CH_2$	0.8
85	T	85	O	Et	H	NO_2	NO_2	$C(Me)_2CH_2$	0.8
86	T	86	O	*i*Pr	H	NO_2	NO_2	$C(Me)_2CH_2$	0.8
87	T	87	O	H	H	NO_2	NO_2	$C(Me)_2CH_2$	0.8
88	T	88	O	OMe	H	NO_2	NO_2	$C(Me)_2CH_2$	0.2
89	T	89	O	Cl	H	NO_2	NO_2	$C(Me)_2CH_2$	0.2
90	T	90	O	Br	H	NO_2	NO_2	$C(Me)_2CH_2$	0.2
91	I	91	H	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
92	I	92	H	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
93	I	93	Me	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
94	I	94	Me	Et	H	H	RCO	$C(Me)_2CHMe$	0.8
95	I	95	Et	Me	H	H	RCO	$C(Me)_2CHMe$	0.8
96	I	96	Me	Et	H	H	RCO	$C(Me)_2CH_2$	0.8
97	I	97	Et	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
98	I	98	Me	H	H	H	RCO	$C(Me)_2CH_2$	0.2
99	I	99	Me	*i*Pr	H	H	RCO	$C(Me)_2CH_2$	0.2
100	I	100	Me	Et	H	H	RCO	$C(Me)_2CHEt$	0.8
101	I	101	Me	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
102	I	102	Me	Me	H	H	RCO	$C(Me)_2CHEt$	0.8
103	I	103	Me	H	H	H	RCO	$C(Me)_2CHMe$	0.8
104	I	104	H	Me	H	H	RCO	$C(Me)_2CH_2$	0.8
105	I	105	Me	H	H	H	RCO	$C(Me)_2CH_2$	0.2
106	NB	54	O	H	H	NO_2	NO_2	OMe	0.8
107	NB	55	O	Me	H	NO_2	NO_2	OEt	0.8
108	NB	56	O	Me	H	NO_2	NO_2	OMe	0.8
109	NB	57	O	Et	H	NO_2	NO_2	OEt	0.8
110	NB	58	O	*i*Pr	H	NO_2	NO_2	OMe	0.8
111	NB	59	O	OMe	H	NO_2	NO_2	Br	0.8
112	NB	60	O	OMe	H	NO_2	NO_2	OMe	0.8
113	NB	61	H	H	H	NO_2	RCO	OMe	0.8
114	NB	62	O	OMe	H	*t*Bu	NO_2	H	0.8
115	NB	66	O	OMe	H	NO_2	NO_2	Me	0.2
116	NB	67	Me	H	H	NO_2	RCO	OMe	0.2

Table I. *(Continued.)*

Description[a]	*Family*[b]	*Molecules*	R_1[c]	*Position 2*	*Position 6*	*Position 3*[d]	O_1[e]	*Position 4*[f]	*Odor*[g]
117	NB	68	O	OMe	Me	*t*Bu	NO_2	H	0.2
118	NB	73	O	Me	H	NO_2	NO_2	OMe	0.8
119	NB	74	O	H	H	NO_2	NO_2	OEt	0.8
120	NB	76	O	*t*Bu	H	NO_2	NO_2	OMe	0.2
121	NB	77	O	Br	H	NO_2	NO_2	OMe	0.2
122	NB	78	H	OMe	H	NO_2	RCO	H	0.2

[a]Descriptions 106–122 correspond to a second description of the 17 molecules (54–78) with two orientations. [b](T): tetralin; (I): indan; (NB): nitrobenzene compound. [c]For the NO_2 substituent, oxygen R_1 is considered as sp^3. [d]Mecy represents a -CH_2-cyclic group. [e]O_1 represents the sp^2 oxygen of the RXO function (RCO or NO_2). [f]Substituents or pseudo-substituents obtained after the fictious cutting of T and I respectively. Substituent -$C(Me)_2CH_2$- was assimilated to a *t*Bu, substituents -$C(Me)_2$CHMe-, -$C(Me)_2$CHMecy, and $C(Me)_2$CHEt- were assimilated to a *t*Am, substituent $C(Me)_2C(Me)_2$- to -$C(Me)_2$*t*Bu (see body of text). [g]Coded as 0.8 for musks and 0.2 for non-musks.

Table II. Different Charton effective upsilon (U_x) values representing the steric hindrances of the different substituents.

Substituent	U^a *(upsilon)*	*Value*
H	U_{ef}	0
Me/Mecy	U_{ef}/U_{est}	0.52
Et	U_{est}	0.56
*i*Pr	U_{ef}	0.76
*t*Bu	U_{ef}	1.24
*t*Am	U_{ef}	1.63
$C(Me)_2$*t*Bu	U_{ef}	2.43
OH	U_{ef}	0.32
OMe	U_{est}	0.42
OEt	U_{ef}	0.48
CHO	U_{min}	0.5
COMe	U_{min}	0.5
CH_2OMe	U_{ef}	0.63
O(sp^3)NO_2	U_{est}	0.4
O(sp^2)NO_2	U_{est}	0.4
O(sp^2)RCO	U_{est}	0.45
F	U_{ef}	0.27
Cl	U_{ef}	0.55
Br	U_{ef}	0.65
CN	U_{ef}	0.4
NO_2	U_{min}	0.35

[a]U_{ef} represents the effective upsilon value given by Charton. U_{min} represents the minimum upsilon value given by Charton. U_{est} represents our estimation of the upsilon value.

fied because their second descriptions were correctly classified.

Only nitrobenzene compounds 60 and 69 were incorrectly classified. Non-musky compound 69 with only one description had a methoxy group located in position 6 between the *t*-butyl and NO_2 groups. Molecule 60 was a musk but its two descriptions (60 and 112) were incorrectly classified. It had two methoxy groups lying between two bulky substituents in position 1, 3 and 5 (fig 1a). It seems proper here to consider that a methoxy group between two bulky groups transmits steric effects and significantly modifies the angle between the plane of the nitro group and the plane of the benzene nucleus, which is assumed to be important in hydrogen bonding [4]. When an NN with six neurons on the hidden layer (8-6-1) was used to increase the descriptor/weight ratio [26], the classification rate was almost unchanged (97.1 *vs* 98.1%).

The discriminant analysis of the 122 descriptions of sample A yielded a result of 65.7% (69/105) of correctly classified molecules. Other linear discrimination methods such as PCA, CFA, and CSA [27, 28] gave rather poor results.

BL reduced samples

The number of incorrectly classified molecules varied between zero and four per sample, *ie* a correct classification rate of 96.4–98.6% for all the 11 BL samples.

Prediction on the BT test samples

The network with eight neurons (8-8-1) on the input layer yielded a result of 21 incorrectly predicted molecules for the 11 BT test samples (9 tetralins and 12 nitrobenzene compounds), *ie* 80% (84/105) of correctly predicted molecules (table III). The predic-

Table III. Results obtained in the classification phase carried out on sample A and those obtained in the prediction phase carried out on the BT test samples.

Families[a]	*Classification phase*		*Prediction phase*	
	Incorrectly classified molecules		*Incorrectly predicted molecules*	
	Musks	*Non-musks*	*Musks*	*Non-musks*
Tetralins	–	–	13, 16, 17, 48	9, 15, 43, 47, 49
Nitrobenzenes	60	69	59, 60, 61, 62, 64	66, 67, 68, 70, 72, 76, 77

[a]There are no indans in the table because they were always correctly classified and predicted.

tion rate was almost unchanged (80%) when an NN with six neurons on the hidden layer (8-6-1) was used.

A network with nine descripteurs (9-9-1) on the input layer (addition of an aromatic nucleus steric descriptor in position 5 (fig 1a)) was used to assess the influence of this position on the results obtained. The network yielded 99% (104/105) of correctly classified molecules for sample A and a mean of 96.6–98.8% of correctly classified molecules for the whole of the BL samples (network with eight neurons: sample A: 98.1% (103/105); BL samples: 96.4–98.6%). However, only 70.5% of all molecules in the BT test samples were correctly predicted by the network with nine neurons on the input layer compared with the 80% of network with eight neurons.

This result suggests that the position 5 substituent is not very important and provides further justification for our not taking it into consideration.

Contribution of the different descriptors

The contribution of each descriptor to the classification was measured using the strategy outlined by Chastrette and Zakarya [20]. All input values corresponding to the descriptor under evaluation were taken as zero. The results of the eight computations made by successively removing one of the eight descriptors show that the relative importance of the different descriptors varied in the following order: $D_3 = D_6 < D_8 < D_1 < D_7 < D_4 < Do_1 < D_2$.

The most important descriptors contributing to the description were as follows: the steric hindrance descriptor of R_2 located alpha to the RXO group (D_2); the descriptors of the oxygen sp^2 of the RXO group (Do_1); and the descriptor of the position 4 (D_4) substituent alpha to the tertiobutyl or assimilated group. This result is in agreement with the order of influences of the different descriptors obtained by Chastrette and Zakarya [20] in their study of a sample comprising only carbonyl tetralins and indans. Descriptor (Do_1) of the RXO group in this study has a very important role considering that it models the difference between the NO_2 and RCO groups. Chastrette and Zakarya [4] assumed that RCO and NO_2 groups accept hydrogen bonds with the same angle of 135° with respect to the axis of the NO_2 or C=O groups. However, the optimal angle between the plane of the functional group and the plane of the nucleus is no doubt different for these two groups and the hindrance of the neighboring substituents is of a different type. Our results show that the network is capable of taking this difference into consideration even if not much is known about the nature of the difference at present.

Table IV. Contribution of the different steric and electronic descriptors to the classification of sample A.

Descriptor[a]	*Percentage of incorrectly classified molecules*
D_1	22.9
D_2	49.5
D_3	13.3
D_4	38.1
Do_1	46.7
D_6	13.3
D_7	31.4
D_8	18.1

[a]D_1, D_2, D_3, D_4 and D_6 correspond respectively to the various steric descriptors indicated as a subscript. Do_1 represents the steric hindrance descriptor of RCO and NO_2 oxygen sp^2 (O_1). D_7 and D_8 represent electronegativity descriptors at positions 2 and 1 (R_2 and R_1).

Conclusion

We have shown in the present study that it is possible to model the structure–odor relationship of musk for a sample of 105 molecules comprising tetralins, indans, and nitrobenzene compounds. A three-layer back-propagation NN was used to obtain a correct-classifi-

cation rate of 98.1% (compared with 65.7% obtained by discriminant analysis) and a correct prediction rate of 80%.

The structure was described from a simple common underlying skeleton and the substituents were coded with eight descriptors, six steric hindrance descriptors and two electronegativity descriptors. The analysis of the contribution of each descriptor to the classification revealed the importance of the steric hindrance in the *ortho* position of an RXO group (X = O or N), which agrees with results obtained in previous studies.

The present study shows that it is possible to combine the three main families of musks under a unique family in which the RCO and NO_2 substituents of the chosen underlying skeleton are considered as equivalents in terms of their interaction through hydrogen bonding with proteinic olfactory receptors. Furthermore, it suggests that the use of NN can be extended to the study of other sets of odorant compounds.

Acknowledgment

We thank the Roudnitska Foundation for the financial support to CE.

References

1 Chastrette M, Zakarya D, Elmouaffek A (1986) *Eur J Med Chem* 21, 505–510
2 Narvaez JB, Lavine BK, Jurs PC (1986) *Chem Senses* 11, 145–156
3 Bersuker IB, Dimoglo AS, Gorbatchov MY, Vlad PF, Pesaro M (1991) *New J Chem* 15, 307–320
4 Chastrette M, Zakarya D (1988) *CR Acad Sci Paris Ser II* 307, 1185–1188
5 Aoyama T, Ichikawa H (1991) *Chem Pharm Bull* 39, 372–378
6 Cherqaoui D, Villemin (1994) *J Chem Soc Faraday Trans* 90, 97–102
7 Villemin D, Cherqaoui D, Cense JM (1993) *J Chim Phys* 90, 1505–1519
8 Bodor N, Harget A, Huang MJ (1991) *J Am Chem Soc* 113, 9480–9483
9 Song XH, Chen Z, Yu RQ (1993) *Computers Chem* 17, 303–308
10 Chang SM, Iwasaki Y, Suzuki M, Tamiya E, Karube I, Muramatsu H (1991) *Anal Chim Acta* 249, 323–329
11 Nakamoto T, Fukuda A, Moriizumi T, Asakura Y (1991) *Sensors Actuat* B3, 221–226
12 Zupan J, Gasteiger J (1991) *Anal Chim Acta* 248, 1–30
13 Song XH, Yu RQ (1993) *Chemometrics Intell Lab Syts* 19, 101–109
14 Jansson PA (1991) *Anal Chem* 63, 357A–362A
15 Smits JMR *et al* (1992) *Anal Chim Acta* 258, 11–25
16 Kosko B (1992) *Neural Network and Fuzzy Systems. A Dynamic Approach to Machine Intelligence*, Prentice-Hall International Editions, London, UK
17 Chastrette M, De Saint-Laumer JY (1991) *Eur J Med Chem* 26, 829–833
18 Rumelhart DE, Hinton CE, Williams RJ (1986) *Nature (Lond)* 323, 533–536
19 Chastrette M, De Saint-Laumer JY, Peyraud JF (1992) *SAR and QSAR in Environmental Research* 1, 221–231
20 Chastrette M, Zakarya D, Peyraud JF (1994) *Eur J Med Chem* 29, 343–348
21 Lefevre J (1976) *Introduction aux Analyses Statistiques Multidimensionnelles*, Masson, Paris, France
22 Dagnelie P (1978-79) *Théorie et Méthodes Statistiques: Applications Agronomiques*. Presses agronomiques de Gembloux, Belgium
23 Charton M (1983) *Top Curr Chem* 114, 57–91
24 Boyd RJ, Boyd SL (1992) *J Am Chem Soc* 114, 1652–1655
25 Pauling L (1960) *The Nature of Chemical Bond and the Structure of Molecules and Crystals*, Third Edition, Cornell Univ Press, NY, USA
26 So S, Graham Richards W (1992) *J Med Chem* 35, 3201–3207
27 McFarland JW, Gans DJ (1986) *J Med Chem* 29, 505–514
28 McFarland JW, Gans DJ (1987) *J Med Chem* 30, 46–49