QSAR Modeling of α-Campholenic Derivatives with Sandalwood Odor

Assia Kovatcheva,[†] Gerhard Buchbauer,*[,†] Alexander Golbraikh,[§] and Peter Wolschann[‡]

Institute of Pharmaceutical Chemistry, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria, and Institute of Theoretical Chemistry and Molecular Biology, University of Vienna, Waehringer Strasse 17, A-1090 Vienna, Austria, and Laboratory of Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, CB 7360, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received August 7, 2002

Three-dimensional quantitative structure−activity relationship (3D-QSAR) models were developed for a series of 44 synthetic α-campholenic derivatives with sandalwood odor. These compounds have complex stereochemistry as they contain up to five chiral atoms. To address stereospecificity of odor intensity, a 3D-QSAR method was developed, which does not require spatial alignment of molecules. In this method, compounds are represented as derivatives of several common structural templates with several substituents, which are numbered according to their relative spatial positions in the molecule. Both wholistic and substituent descriptors calculated with the TSAR software were used as independent variables. Based on published experimental data of sandalwood odor intensities, two discrete scales of the odor intensity with equal or unequal intervals between the threshold values were developed. The data set was divided into a training set of 38 compounds and a test set of six compounds. To build QSAR models, a stepwise multiple linear regression method was used. The best model was obtained using the unequal scale of odor intensity: for the training set, the leave one out cross-validated R^2 (q^2) was 0.80, the correlation coefficient R between actual and predicted odor intensities was 0.93, and the correlation coefficient R_0^2 for the test set was 0.95. The QSAR models developed in this study contribute to the better understanding of structural, electronic, and lipophilic properties responsible for sandalwood odor. Furthermore, the QSAR approach reported herein can be applied to other data sets that include compounds with complex stereochemistry.

INTRODUCTION

The stereochemistry of biologically active substances is an important factor in many QSAR investigations. Amino acids, carbohydrates, and lipids as well as many other natural and artificial receptor ligands are examples of chiral molecules, which have biological properties different from those of their enantiomers.[1,2] Physiological effects of many drugs depend strongly on their stereochemistry,[3−5] L- and D-enantiomers of amino acids have different taste,[6,7] and many insecticides are stereospecific.[8] According to the phenomenon of stereospecificity, only one of the enantiomers of any given compound manifests the desired biological activity.

A human organism is capable of distinctly recognizing enantiomers.[9] Fragrance compounds, which are of particular interest in the perfume industry and complementary medicine are examples of the influence of chirality on the biological effect,[10] such as olfaction and odor impression and intensity.[11−14]

Up to now, the mechanisms of olfaction are not completely understood. Buck and Axel[15] showed that odorant molecules activate a specific neuronal heterotrimeric G-protein coupled receptor, which increases the concentration of intracellular cyclic adenosine 3′,5′-monophosphate (cAMP) and subsequently leads to the opening of cyclic nucleotide-gated channels (CNG) in the plasma membrane.[16] It has been demonstrated that odor-adaptation is related to the permeability of CNG and particularly CNGA4 channel subunit for Ca^{2+}/Calmodulin.[17,18] Another line of investigations dealt with the discriminatory ability of the olfactory system.[19,20] A combinatorial receptor coding scheme[21] was proposed to encode odor identities. This scheme was validated by a stereotyped sensory map in the olfactory cortex.[22]

East Indian sandalwood oil is a valuable constituent of perfume compositions. A large number of natural as well as artificial compounds with this typical odor tonality is known already. Among all diverse substitutes, a group of α-campholenic derivatives exhibits an odor most similar to the natural sandalwood.[23,24] On the basis of the combinatorial receptor coding scheme, Bajgrowicz et al.[25] identified similar chemical features among compounds exhibiting the same type of odor and synthesized two α-campholenic aldehyde derivatives (Table 1, compounds **1**−**4**, **5**−**8**). The odor impression of their optically pure enantiomers were evaluated by taking into account both qualitative properties and quantitative thresholds. The enantiomer with the highest intensity known as Javanol has a strong sandalwood-like odor.

Madrol (Table 1, compounds **9**, **10**) is another α-campholenic aldehyde derivative with sandalwood odor, which was synthesized and described by various groups.[26−28]

* Corresponding author phone: 0043 1 4277 55150; fax: 0043 1 4277 9551; e-mail: gerhard.buchbauer@univie.ac.at.

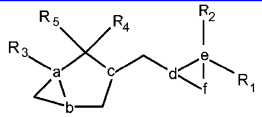
[†] Institute of Pharmaceutical Chemistry, University of Vienna.

[‡] Institute of Theoretical Chemistry and Molecular Biology, University of Vienna.

[§] University of North Carolina at Chapel Hill.

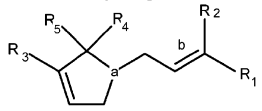
10.1021/ci020296n CCC: $25.00
Published on Web 01/07/2003

Table 1. Data Set of 44 α-Campholenic Derivatives

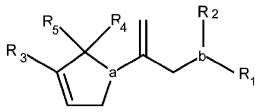


Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	c	d	e	Scale1	Scale2	Ref.
1	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	R	R	S	S	5	0.95	25
2	-H*	-H*	$-CH_3$	$-CH_3$	$-CH_3$	S	R	R	R	R	0	0	25
3	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	S	S	S	S	4	0.80	25
4	$-CH_3$	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	R	S	S	R	R	0	0	25
5	$-CH_2OH$	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	R	R	S	S	5	0.95	25
6	-H*	-H*	$-CH_3$	$-CH_3$	$-CH_3$	S	R	R	R	R	1	0.05	25
7	$-CH_2OH$	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	S	S	S	S	2	0.20	25
8	$-CH_2OH$	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	S	S	R	R	3	0.50	25

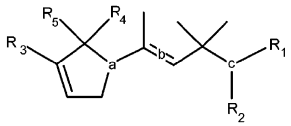
*Substituents R_1 and R_2 are attached to atom f.



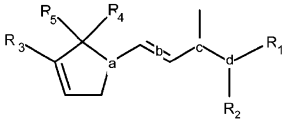
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	Scale1	Scale2	Ref.
9	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	4	0.80	28
10	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	3	0.50	28
11	$-CH_2OH$	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	5	0.95	28
12	$-CH_2OH$	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	Z	0	0	28
13	$-CH_2OH$	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	5	0.95	28
14	$-CH_2OH$	$-CH_2CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	4	0.8	28



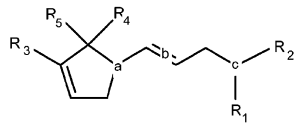
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	Scale1	Scale2	Ref.
15	-CHO	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	R	3	0.50	40
16	-CHO	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	S	3	0.50	40
17	-CHO	-H	$-CH_3$	$-CH_3$	$-CH_3$	R	-	0	0	40
18	-CHO	-H	$-CH_3$	$-CH_3$	$-CH_3$	S	-	0	0	40
19	$-CH_3$	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	R	R	4	0.80	40
20	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	S	3	0.50	40
21	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	S	2	0.20	40
22	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	R	2	0.20	40



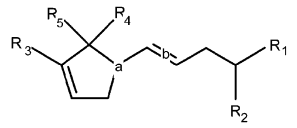
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	c	Scale1	Scale2	Ref.
23	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	R	2	0.20	59
24	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	S	3	0.50	59
25	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	R	3	0.50	59
26	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	S	4	0.80	59



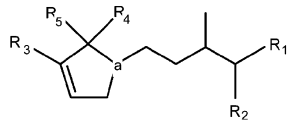
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	c	d	Scale1	Scale2	Ref.
27	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	S	R	4	0.80	60
28	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	R	S	0	0	60
29	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	R	S	0	0	60
30	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	S	R	5	0.95	60
31	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	R	R	0	0	60
32	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	S	S	0	0	60
33	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	S	S	1	0.05	60
34	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	E	R	R	2	0.2	60



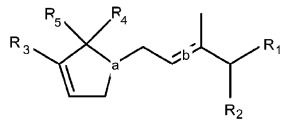
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	c	Scale1	Scale2	Ref.
35	-OH	$-CH_2CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	R	2	0.20	61
36	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	R	1	0.5	61



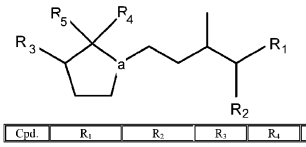
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	Scale1	Scale2	Ref.
37	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	3	0.05	61



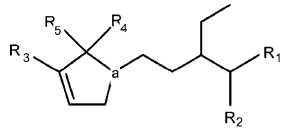
Cpd.	R_1	R_2	R_3	R_4	R_5	a	Scale1	Scale2	Ref.
38	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	S	3	0.50	62
39	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	3	0.50	62



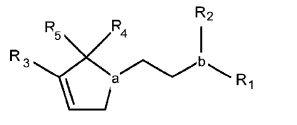
Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	Scale1	Scale2	Ref.
40	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	E	4	0.80	61



Cpd.	R_1	R_2	R_3	R_4	R_5	a	Scale1	Scale2	Ref.
41	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	0	0	62



Cpd.	R_1	R_2	R_3	R_4	R_5	a	Scale1	Scale2	Ref.
42	-OH	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	2	0.20	62



Cpd.	R_1	R_2	R_3	R_4	R_5	a	b	Scale1	Scale2	Ref.
43	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	$-CH_3$	R	R	4	0.80	27
44	$-CH_3$	$-CH_2OH$	$-CH_3$	$-CH_3$	$-CH_3$	S	R	3	0.50	27

Applying enantioselective gas chromatography separation, Mosandl et al.[29] found that its (R)−(+)-enantiomer exhibits a strong sandalwood odor in contrast with the (S)−(−)-enantiomer with a faint milky odor.

An ongoing interest in the development of new compounds with the sandalwood odor and a challenging task to understand the role of enantioselectivity in the odor intensity have provided a strong incentive for Quantitative Structure−Activity (QSAR) studies of these compounds. The QSAR studies of a highly diverse set of compounds with sandalwood odor were reported recently[30] using the Minimum Topological Difference (MTD) method.[31] This approach includes conformational analysis and spatial alignment of molecules. Due to the high diversity of the data set, the authors were unable to superimpose all molecules, which is a typical limitation of many 3D-QSAR studies. Consequently, the resulting model was unable to correctly predict the odor threshold of a large fraction of compounds, which were not structurally aligned with the rest of the data set. Several 3D-QSAR alignment-independent QSAR approaches have been developed to circumvent the alignment problem. Examples include the approach using autocorrelation functions (Broto,[32] Gasteiger et al.,[33,34] and Clementi et al.[35]), WHIM descriptors (Todeschini and Gramatica[36]), and comparative molecular moment analysis (Silverman and Platt[37]). One of the most promising alignment-independent methods based on grid-independent descriptors has been developed recently (Pastor and Cruciani[38]).

Herein, we introduce a novel 3D-QSAR approach that allows the description and comparison of chemical structures without performing their direct alignment. This method is particularly suitable for studies of molecules with complex

Table 2. Scales of Odor Intensities

	intensity levels	
odor intensity	scale 1	scale 2
odorless	1	0
very weak	2	0.05
weak	3	0.20
average	4	0.50
strong	5	0.80
very strong	6	0.95

stereochemistry. We apply this approach to a series of α-campholenic derivatives with known configurations of their chiral atoms. All compounds were represented as derivatives of several common structural templates with five substituents and numbered according to their relative spatial positions. The QSAR model was developed using Multiple Linear Regression (MLR) and validated using an external test set of compounds with known odor thresholds. The QSAR approach described in this paper can be applied to other data sets with complex stereochemistry.

MATERIALS AND METHODS

Data Set and Measures of Odor Intensity. The structures and odor properties of 44 α-campholenic aldehyde derivatives selected for QSAR studies from several publications are given in Table 1. The main criteria for including a compound in the data set were as follows: (i) the presence of at least one chiral atom and (ii) availability of both qualitative and quantitative data on the sandalwood scent. The quantitative description of odor is a daunting task. Often, it is supported by additional qualitative description, which is required to be very accurate and "as scientific as the knowledge about the chemical structure."[39] A typical sandalwood odor is described using qualitative descriptors such as strong, woody, warm, creamy, milky, sweet, etc.

In some cases, the quantitative description of the scent intensity, so-called odor threshold, is given by comparing the intensities of individual qualitative properties. For example, compound **19** (Table 1) exhibits elegant and powerful sandalwood odor, whereas compounds **21** and **22** are less rich in the precious "lait de santal".[40] In ref 25, the odor threshold was defined in ng/L using a gas chromatography sniffing technique. In ref 28 compounds **9**−**14** were ranked by seven panelists. Their odor thresholds were obtained using the so-called triangle method[41] and expressed in mol/L. In addition to the odor perception, many odorants also stimulate trigeminal receptors. Sandalwood odorants do not cause the trigeminal sensation,[42] which allows better quantitative description of the odor thresholds.[39]

Although different authors define odor intensities differently, both quantitative and qualitative odor descriptions were used to assign the value of the intensity for each compound. We have transformed published experimental odor thresholds and descriptions into two scales (Table 2) of sandalwood odor intensities. Scale 1 was characterized by equal intervals between odor intensity levels ranging from zero to five. With this scale, the data set was divided into six groups of compounds characterized by different values of the odor intensity and the markedly biggest group included compounds with the levels of odor intensity close to the average level.

To achieve a more even distribution of all compounds within the whole interval of odor intensities, we have also introduced scale 2 with six unequal intervals between intensity levels; this scale provided a better separation between the average level of 0.5 and the values of weak (0.2) and strong (0.8) intensities. The quality of the experimental data was insufficient to discern substances with the maximal intensity of the sandalwood odor since different methods of defining the odor thresholds were used. Thus, the odor intensity of one was introduced only as an upper threshold of the intensity, but none of the compound was actually assigned this value (cf. Table 1).

Molecular Representation and Descriptors. Three-dimensional structures of all molecules were generated, and their geometries were optimized using Sybyl 6.5.[43] Molecular mechanics calculations were performed using the Tripos force field with Gasteiger−Hückel charges. The calculations were terminated, if the energy difference or the energy gradient were smaller than 0.01 and 0.001 kcal/mol, respectively. The resulting 3D-structures were then exported to TSAR 3.21 software.[44]

Due to the complex stereochemistry of α-campholenic derivatives the most important step, which afforded the QSAR modeling, was to establish their unique representation. To this end, we have introduced a chemical encoding scheme according to which each molecule was described as a template with a defined number of substituents attached to this template by a single bond. A single hydrogen atom may also serve as a substituent. All molecules had the same number of substituents (Table 1). Certainly, there exist several ways to represent molecules as template with different substituents. After some experimenting, we chose an encoding scheme where the number of substituents was five. These substituents were assigned as shown in Table 1.

All substituents were numbered according to their positions in molecules. The consistent numbering of substituents is very important. Particular care should be taken, if two substituents are attached to the same atom of the template. Consider molecules in Table 1. All molecules have methyl groups in positions R4 and R5; however, these methyl groups are not equivalent. They are mirror images of each other in the plane defined by the central atom they are connected to and two adjacent carbon atoms in the five-membered ring. It is important to decide as to which of the two methyl groups is R4 and which one is R5. The plane described above serves this purpose as follows. We can define the atom, to which substituent R3 is attached as having higher priority than the atom designated as *a* (i.e. to which a long chain is attached) in Table 1. Then the side of the plane, in which the atom with the lower priority is positioned anticlockwise with respect to the atom with the higher priority around the central atom, is defined as positive. The other side of the plane is defined as negative. The R4 substituent can be defined as located on the positive side of the plane, and R5 can be defined as located on the negative side (Figure 1). Due to this difference of the plane sides, methyl groups R4 and R5 have different properties, and in stereochemistry they are referred to as enantiotopic groups.

Substituents R1 and R2 (Table 1) are attached to the same carbon atom in all molecules. For compound **1**, groups CH_2-OH and CH_3 were designated as R1 and R2, respectively (see Table 1). In molecules **9**−**44,** substituents R1 and R2

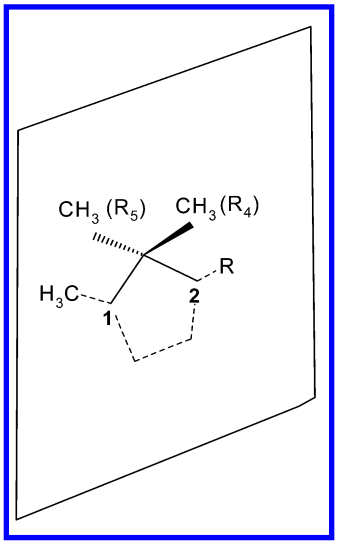


Figure 1. Enantiotopic methyl groups R4 and R5 attached to the cyclopentane ring. Atom 1 has higher priority than atom 2.

are connected to the distal branching point of the acyclic moiety. As soon as methyl groups attached to the five-membered ring were correctly assigned to R3, R4, and R5, it was not very difficult to decide which groups of the molecule corresponded to substituents R1 and R2 of molecule **1.** For example, it could be done by placing a molecule in such a way that the spatial orientations of methyl groups of the lipophilic moiety had the same orientations as those of molecule **1**. Thus, it was found that in molecule **2**, two hydrogen atoms shown in Figure 2 corresponded to substituents R1 and R2 of molecule **1** (Figure 2). This comparison was performed for all molecules.

Molecule **1** was used as a template to assign substituents R1 and R2 for the following reasons. (i) This molecule has very strong odor. (ii) The information about the odor properties is described in detail.[25] (iii) It has five known chiral atoms and is more rigid than the majority of other compounds of the data set.

Molecular descriptors were calculated with the TSAR 3.21 software.[44] Descriptors were obtained for both the entire molecule and substituents defined. TSAR affords calculation of the following descriptors: Molecular Surface Area and Volume, Moments of Inertia, Ellipsoidal Volume, Verloop Parameters,[45,46] Dipole Moments, Lipole Moments, Molecular Mass, Wiener Index,[47] Molecular Connectivity Indices,[48,49] Molecular Shape Indices,[50] Electrotopological State Indices,[51] LogP, Number of defined atoms (carbons, oxygens etc.), and Number of defined groups (methyl, hydroxyl etc.).

The majority of these descriptors can be calculated for both the entire molecule (wholistic descriptors) and substituents. There are also descriptors, which can be calculated for substituents only. They include Bond Dipole and Bond Lipole descriptors.[44] Due to our definition and numbering of substituents, these descriptors are capable of distinguishing between stereoisomers.

Descriptors with the same values for all compounds (Table 1) were discarded. Pairwise correlation analysis of the remaining descriptors was performed. For each pair of descriptors, the correlation coefficient was calculated. If for two descriptors the correlation coefficient was higher than 0.65, regressions were built using descriptor subsets containing only one of these highly correlated descriptors. The descriptor with lower *t*-value[44] was discarded.

Training and Test Set. The data set was randomly divided into the training and test set of 38 and six compounds, respectively. The test set included compounds 4, 13, 22, 28, 34, and 40 (Table 1). All other compounds were included into the training set.

QSAR Model Development and Validation. To develop QSAR models, stepwise MLR analysis with leave-one-out (LOO) cross-validation was applied to the training set. *F*-to-enter and *F*-to-leave values were 3 and 2, respectively. At first, multiple models have been built using large subsets of randomly selected descriptors. Then the number of descriptors included in the procedure was gradually reduced by the selection of those, which had higher *t*-values[44] or which appeared with higher frequency in previous models. Models with the number of descriptors no higher than eight, cross-validated $R^2(q^2)$ greater than 0.6, *F*-ratio higher than 20, and correlation coefficient *R* higher than 0.8 between predicted and observed odor intensities were validated using compounds of the test set. The following statistical characteristics of the models were used: correlation coefficient *R* and coefficients of determination R_0^2 and $R_0'^2$ between predicted and observed intensities and slopes *k* and *k*′ of regressions through the origin of predicted versus observed and observed versus predicted intensities.[52] Models were considered to have high predictive ability, if $R^2 > 0.6$, $(R^2 - R_0^2)/R^2 < 0.1$ or $(R^2 - R_0'^2)/R^2 < 0.1$, and the corresponding $0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$. Preferably, both R_0^2 and $R_0'^2$ had to be close to each other.

RESULTS AND DISCUSSION

For a set of 44 α-campholenic derivatives with known configuration of at least one chiral center QSAR models were built using stepwise MLR analysis. After calculation of descriptors for the entire molecule and for number of substituents ranging between 2 and 6, the best QSAR models were built describing the entire molecule with five substituents presented in Table 1. In the beginning, the training and test set included 38 and six compounds, respectively

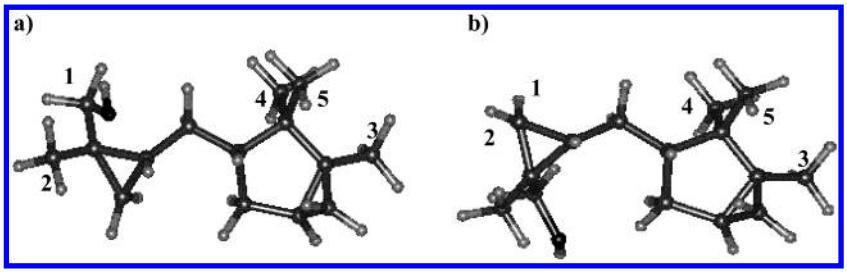


Figure 2. Compounds 1 (a) and 2 (b) and their substituents.

Table 3: Models Developed for a Training Set of 35 Compounds and a Test Set of 6 Compounds

model no.	scale	no. of descriptors	$q(cv)^2$	R	F	s	R_0^2 (test set)	$R_0'^2$ (test set)	k
1	2	8	0.82	0.96	42.71	0.10	0.44	0.33	0.98
2	2	7	0.74	0.94	28.58	0.88	0.48	0.43	0.96
3	1	6	0.89	0.94	36.43	0.59	0.86	0.86	0.90
4	2	6	0.80	0.93	28.99	0.13	0.95	0.94	1.09
5	2	5	0.71	0.90	23.55	0.16	0.44	0.59	0.87

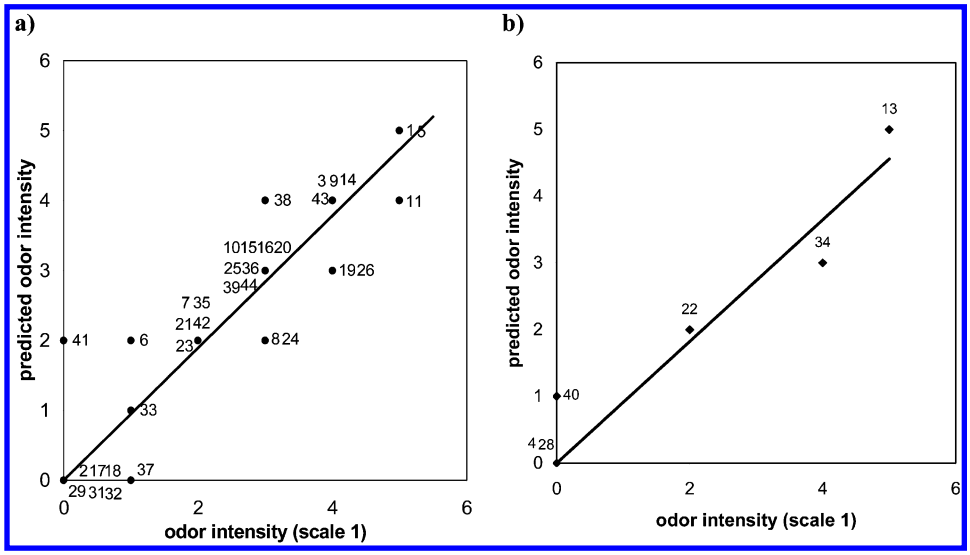


Figure 3. Prediction of odor intensity using scale 1 (Model 3, Table 3): (a) compounds of the training set and (b) compounds of the test set.

Table 4. Descriptors Included in the Best Model Obtained Using Scale 1 (Model 3 in Table 3)

descriptor	coefficient	t-value	t-probability	SE
X_1 molecular volume (subst. 3)	−0.68	−4.11	3.15e-05	0.17
X_2 inertia moment 1 size (subst. 4)	−720.54	−7.77	0	92.70
X_3 dipole moment Z − component (subst. 2)	−5.75	−9.14	0	0.63
X_4 lipole X − component (subst. 5)	1.61	7.24	0	0.67
X_5 bond lipole (subst. 2)	0.70	8.79	0	0.18
X_6 group count for methyl (whole molecule)	243.63	4.71	6.2e-05	0.15

(see the previous section). In the majority of regressions built with different subsets of descriptors, compounds 12, 27, and 30 (Table 1) turned out to be outliers. Every time, after excluding these compounds from the training set, the statistical characteristics such as LOO q^2 and correlation coefficient R for the training set were improving significantly. Therefore, these compounds were eventually excluded from the training set, and QSAR models were developed for the remaining 35 compounds using both scales of odor intensities (Table 2). Several QSAR models were obtained, which had excellent q^2 and R values for the training set (Table 3). These models were validated using the compounds of the test set. Models 1, 2, and 5 (Table 3) did not demonstrate high predictive power: the correlation coefficient R^2 between predicted and observed odor intensity for the test set did not exceed 0.44.

The best QSAR model obtained using scale 1 (Model 3, Table 3) can be represented by the following equation

$$\text{od. int.} = -0.68*X_1 - 720.54*X_2 - 5.75*X_3 + 4.88*X_4 + 1.61*X_5 + 0.70*X_6 + 243.63 \quad (1)$$

$$s = 0.59,\ q^2 = 0.89,\ R = 0.94, \text{ and } F = 36$$

where s is the standard error of prediction and F is the F-ratio.

Predictive ability of this model is characterized by the following statistics: $R = 0.94$, $R_0^2 = 0.86$, $R_0'^2 = 0.86$, $k = 0.90$, $k' = 1.02$. Thus this model satisfied all the criteria of a predictive model.[52] The correlation between observed and predicted odor intensity for this model is shown in Figure 3. The descriptors and their characteristics are given in Table 4.

The best QSAR model obtained using scale 2 can be represented by the following equation (Model 4, Table 3)

$$\text{od. int.} = -0.13*X_1 - 167.85*X_2 - 1.06*X_3 + 1.07*X_4 + 0.27*X_5 + 0.14*X_6 + 56.32 \quad (2)$$

$$s = 0.13,\ q^2 = 0.80,\ R = 0.93, \text{ and } F = 29$$

Predictive ability of this model is characterized by the following statistics: $R = 0.93$, $R_0^2 = 0.95$, $R_0'^2 = 0.94$, $k = 1.09$, $k' = 0.88$. Thus this model satisfied all the criteria of a predictive model.[52] High predictive power of this model is demonstrated in Figure 4. The descriptors and their characteristics are given in Table 5.

The model obtained with odor intensities assigned according to scale 2 showed better predictive ability. For two compounds of the test set, scale 1 levels of odor intensity closest to the predicted values are different from the observed

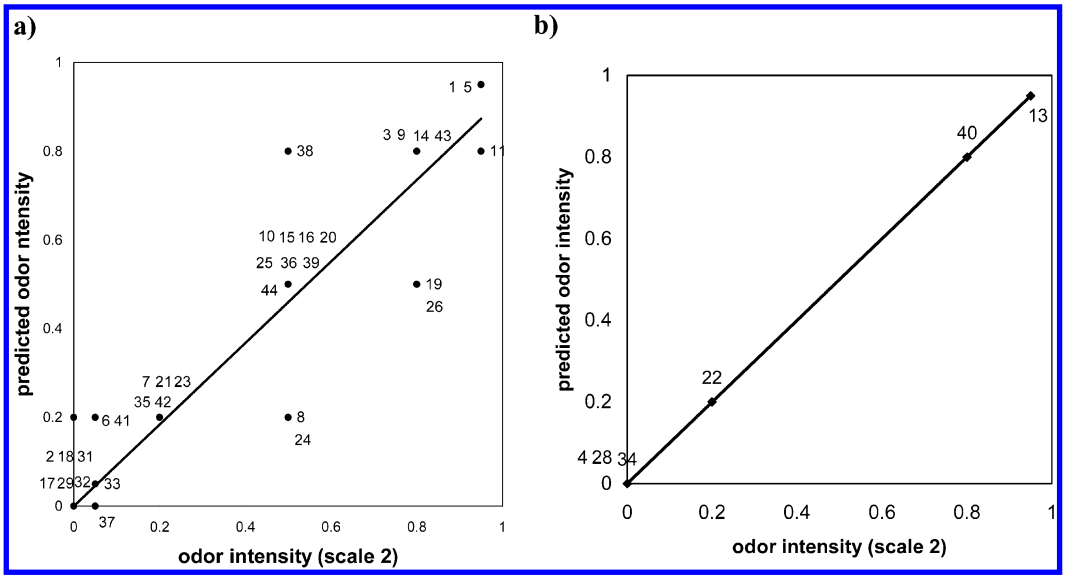


Figure 4. Prediction of odor intensity using scale 2 (Model 4, Table 3): (a) compounds of the training set and (b) compounds of the test set.

Table 5. Descriptors Included in the Best Model Obtained Using Scale 2 (Model 4 in Table 3)

descriptor	coefficient	t-value	t-probability	SE
X_1 molecular volume (subst. 3)	−0.13	−3.45	1.8e-03	0.04
X_2 inertia moment 1 size (subst. 4)	−167.85	−7.97	1.12e-08	21.05
X_3 dipole moment Z − component (subst. 2)	−1.06	−7.39	4.75e-08	0.14
X_4 lipole X − component (subst. 5)	1.07	7.02	1.24e-07	0.15
X_5 bond lipole (subst. 2)	0.27	6.46	5.32e-07	0.04
X_6 group count for methyl (whole molecule)	0.13	4.04	0.4e-03	0.03

values assigned according to this scale (Figure 3b). For all compounds of the test set, scale 2 levels of odor intensity closest to the predicted values are equal to the observed values assigned according to this scale (Figure 4b). Thus, the unequal intervals of scale 2 turned out to be more appropriate for defining odor intensities. It affords the distribution of the compounds more similar to their natural distribution by odor intensities.

Many studies on sandalwood odorants show that only certain parts of the molecules are responsible for the expression of their biological activity.[53−57] The best QSAR model built with our approach can be used in the fragrance chemistry to discern the odor properties of new sandalwood odorants. Our model does not include explicit descriptors for substituent R1, which in the majority of molecules contains the functional hydroxyl group. This may imply that the hydroxyl group is essential for the expression of odor properties only in combination with the steric and electrostatic parameters of the lipophilic part of the molecule. Substituent R2 was represented by two descriptors of charge and lipophilicity distribution: Dipole Moment Z-Component and Bond Lipole, respectively. These descriptors also confirm the importance of the interaction between substituents R1 and R2, one of which includes a hydroxyl group, which is important for the expression of sandalwood odor.[58] In the best model, substituents R3, R4, and R5 are represented by Molecular Volume for substituent R3, Inertia Moment 1 Size for substituent R4, and Lipole X-Component for substituents R5. Taking into account that these three substituents are located close to each other, it is obvious that bulky properties of these substituents and their precise location in the lipophilic core are necessary for compounds to exhibit the sandalwood fragrance. The sixth descriptor (Group count for Methyl) shows that in general the number of methyl groups in the whole structure is important.

CONCLUSIONS

The goal of this study was to develop a QSAR model for prediction of odor intensities of α-campholenic derivatives taking into account their chiral properties. Based on the thorough analysis of the experimental data of the sandalwood odor intensity, two scales of the target property with equal or unequal intervals were introduced. A novel procedure based on the determination of spatial orientations of enantiotopic groups was developed to define and number substituents at critical positions of chemical structures, which were used to calculate descriptors. It was shown that a subset of compounds which have identical chemical structures but differ only in the configuration of their chiral centers and, consequently, in their biological activities can be described and distinguished by descriptors of the entire molecule and its substituents. A highly predictive QSAR model has been obtained using the unequal scale. It was validated using external test set of compounds. The model can be used for the design and discovery of new chemicals with the sandalwood odor. The method used to assign substituents can be applied to other data sets with complex stereochemistry. The model contributes also to the better understanding of structural, electronic, and lipophilic properties responsible for sandalwood odor. We found that the interaction of the functional methyl group and the neighboring lipophiling substituent is important for the expression of sandalwood odor. We concluded also that the precise positions of methyl

groups attached to the lipophilic five-member ring also contribute to the activity of α-campholenic derivatives.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Dr. Alexander Tropsha for his attention to this work and valuable suggestions. To Dragoco company, Vienna, we owe thanks for the interest in this study. This study was supported by the DOC-program of the Austrian Academy of Science.

REFERENCES AND NOTES

(1) Horton, H. R.; Moran, L. A.; Ochs, R. S.; Rawn, J. D.; Scrimgeour, K. G. In *Principles of Biochemistry*; Neil Patterson Publishers Prentice Hall: Englewood Cliffs, NJ, 2002.
(2) Potapov, V. M. *Stereochemistry*; Khimia: Moscow, 1988.
(3) DeCamp, W. H. The FDA Perspective on the development of stereoisomers. *Chirality* **1989**, *1*, 2−6.
(4) Hutt, A. J.; Tan, S. C. Drug chirality and its clinical significance. *Drugs* **1996**, *52*, 1−12.
(5) Wendt, S.; Zwingenberger, K. Thalidomide's chirality. *Nature* **1997**, *385*, 303−304.
(6) Solms, J.; Vuataz, L.; Egli, R. H. The taste of L- and D-amino acids. *Experientia* **1965**, *21*, 692−694.
(7) Schiffman, S. S.; Clark, T. B. Gagnon, J. 3d Influence of chirality of amino acids on the growth of perceived taste intensity with concentration. *Physiol. Behav.* **1982**, *28*, 457−465.
(8) Kurihara, N.; Miyamoto, J.; Paulson, G. D.; Zeeh, B.; Skidmore, M. W.; Hollingworth, R. M.; Kuiper, H. A. IUPAC reports on pesticides 37. Chirality in synthetic agrochemicals: Bioactivity and safety consideration. *Pure Appl. Chem.* **1997**, *69*, 1335−1348.
(9) Holmstedt, B.; Frank, H.; Testa, B. *Chirality and Biological Activity*; Alan R. Liss.: New York, 1990.
(10) Chapuis, Ch.; Jacoby, D. Catalysis in the preparation of fragrances and flavor. *Applied Catalysis A: General* **2001**, *221*, 93−117.
(11) Theimer, E. T.; McDaniel, M. R. Odor and optical activity. *J. Soc. Cosmet. Chem.* **1971**, *22*, 15−26.
(12) Ohloff, G.; Giersch, W.; Schulte-Elte, K. H.; Vial Ch. Stereochemistry-odor relations of 1-decalone derivatives and their oxygen analogues. C. *Helv. Chim. Acta* **1976**, *59*, 1140−57.
(13) Schulte-Elte, K. H.; Giersch, W.; Winter, B.; Pamingle, H.; Ohloff, G. Diastereoselective odor perception of alcohols in the ionone series. *Helv. Chim. Acta* **1985**, *68*, 1961−85.
(14) Snowden, R. L.; Sonnay, P.; Ohloff, G.Stereoselective syntheses of (±)-epi-*β*-santalene and (±)-epi-*β*-santalol. *Helv. Chim. Acta* **1981**, *64*, 25−32.
(15) Buck, L.; Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recoognition. *Cell* **1991**, *65*, 175−187.
(16) Dzeja, C.; Hagen, V.; Kaupp, U. B.; Frings, S·Ca^{2+} permeation in cyclic nucleotide-gated channels. *EMBO J.* **1999**, *18*, 131−144.
(17) Munger, S. D.; Lane, A. P.; Zhong, H.; Leindeers-Zufall, T.; Yau, K. W.; Zufall, F.; Reed, R. R. Central Role of the CNGA4 Channel subunit in Ca^{2+}-Calmodulin-Dependent Odor Adataption. *Science* **2001**, *294*, 2172−2175.
(18) Bradley, J.; Reuter, D.; Frings, S. Facilitaion of Calmodulin-Mediated Odor Adataption by cAMP−Gated Channel Subunits. *Science* **2001**, *294*, 2176−2178.
(19) Laska, M.; Teubner, P. Olfactory discrimination ability of human subjects for 10 pairs of enantiomers. *Chem. Senses* **1999**, *24*, 161−170.
(20) Laska, M. Olfactory discrimination ability for aromatic odorants as a function of oxygen moiety. *Chem. Senses* **2002**, *27*, 23−29.
(21) Malnic, B.; Hirono, J.; Sato, T.; Buck, L. Combinatorial receptor codes for odors. *Cell* **1999**, *96*, 713−723.
(22) Zou, Z.; Horowitz, L.; Montmayeur, J.-P.; Snapper, S.; Buck, L. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature* **2001**, *414*, 173−179.
(23) Schulze, K.; Habermann, A. K.; Uhlig, H.; Weber, L.; Kempe, R. Synthesis, structure, and reactions of α-campholenic epoxides. *Liebigs Ann. Chem.* **1993**, *9*, 987−91.
(24) Brunke, E. J.; Klein, E. Chemistry of sandalwood fragrance. In *Fragrance Chemistry. The Science of the Sense of Smell*; Theimer, E. T., Ed.; Academic Press: New York, 1982.
(25) Bajgrowicz, J. A.; Frater, G. Chiral recognition of sandalwood odorants. *Enantiomer* **2000**, *5*, 225−234.
(26) Buchbauer, G.; Lebada, Ph.; Wiesinger, L.; Weiss-Greiler, P.; Wolschann, P. On the odor of the enantiomers of Madrol. *Chirality* **1997**, *9*, 380−385.
(27) Bajgrowicz, J. A.; Frank, I.; Frater, G.; Hennig, M. Synthesis and structure elucidation of a new potent sandalwood-oil substitute. *Helv. Chim. Acta* **1998**, *81*, 1349−1358.
(28) Aida, T.; Harada, M.; Iwai, H.; Amano, A.; Yamasaki, T.; Yamamoto, T. (E)−(R)-2-alkyl-4-(2,2,3-trimethylcyclopent-3-en-1-yl)-2-buten-1-ol, process for preparing the same, and use thereof in perfume compositions. *EP 829463 A2*, **1998**, CAN 128:244210.
(29) Bilke, S.; Mosandl, A. Enantioselective analysis of 2-methyl-4-(2,2,3-trimethylcyclopent-3-en-1-yl)-but-2-enol, 2-methyl-4-(2,2,3-trimethylcyclopent-3-en-1-yl)-but-2-enal and α-campholene aldehyde by capillary gas chromatography. *J. Sep. Sci.* **2001**, *24*(10), 819−822.
(30) Hadaruga, D. I.; Muresan, S.; Bologa, C.; Chiriac, A.; Simon, Z.; Cofar, L.; Naray-Szabo, G. QSAR for cycloaliphatic alcohols with qualitatively defined sandalwood odor characteristics. *Quantitative Structure−Activity Relationships* **1999**, *18*, 253−261.
(31) Simon, Z.; Chiriac, A. M.; Holban, S., Ciubotariu, D. and Mihalas, G. I. *Minimum steric difference. The MTD Method for QSAR Studies*; Research Studies Press: Letchworth, New York, 1984.
(32) Broto, P.; Moreau, G.; Vandycke, C. Molecular structures: perception, autocorrelation descriptor and SAR studies. Autocorrelation descriptor. *Eur. J. Med. Chem.* **1984**, *19*, 66−70.
(33) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.
(34) Anzali, S.; Barnickel, G.; Krug, M.; Sadowski, J.; Wagener, M.; Gasteiger, J.; Polanski, J. The comparison of geometric and electronic properties of molecular surfaces by neural networks: application to the analysis of corticosteroid-binding globulin activity of steroids. *J. Comput.-Aided. Mol. Des.* **1996**, *10*, 521−534.
(35) Clementi, S.; Cruciani, G.; Riganelli, D.; Valigi, R.; Costantino, G.; Baroni, M.; Wold, S. Autocorrelation as a tool for a congruent description of molecules in 3D-QSAR studies. *Pharm. Pharmacol. Lett.* **1993**, *3*, 5−8.
(36) Todeschini, R.; Gramatica, P. New 3D Molecular Descriptors: The WHIM Theory and QSAR Applications. In *3D QSAR in Drug Design*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer/Escom: Dordrecht, 1998; Vol. 2, pp 355−380.
(37) Silverman, B. D.; Platt, D. E. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J. Med. Chem.* **1996**, *39*, 2129−2140.
(38) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.
(39) Market, Th. Structure Activity Relationships and the Subjectivity of odor sensation. In *Advances in Flavors and Fragrances From the Sensation to the Synthesis*; Karl, A. D., Swift, T. J., Eds.; International Ltd.: Padstow, 2002; pp 3−15.
(40) Chapuis, Ch.; Blanc, P.-A. Campholen aldehyde derivatives and their use in perfumery. *EP 694520 A1*, **1996**, CAN 124:261418.
(41) Acree, T. E.; Nishida, R.; Fukami, H. Odor thresholds of the stereoisomers of methyl jasmonate. *J. Agric. Food Chem.* **1985**, *33*, 425−427.
(42) Laska, M.; Distel, H.; Hudson, R. Trigeminal perception of odorant quality in congenitally anosmic subjects. *Chem. Senses* **1997**, *22*, 447−456.
(43) SYBYL 6.6, Molecular Modeling Software, Tripos Inc. 1999.
(44) TSAR 3.2, Oxford Molecular Limited 1999.
(45) Verloop, A.; Hoogenstraaten, W.; Tipker, J. Development and application of new steric substituent parameters in drug design. *Drug Design* **1976**, *7*, 165−207.
(46) Verloop, A.; Tipker, J. A comparative study of new steric parameters in drug design. *Pharmacochem. Libr.* **1977**, *2*, 63−68.
(47) Wiener, H. Structural Determination of Paraffin Properties. *J. Am. Chem. Soc.* **1947**, *69*, 17.
(48) Randic, M. On Characterization on Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(49) Kier, L. B.; Hall, L. *Molecular Connectivity in Structure−Activity Analysis*; Wiley: New York, 1986.
(50) Kier, L. B. A shape index from molecular graphs. *Quantum-Struct. Act. Relat.* **1985**, *4*, 109−116.
(51) Kier, L. B.; Hall, L. *Molecular Structure Desription The electrotopological state*; Academic Press: 1999.
(52) Golbraikh, A.; Tropsha, A. Beware of q^2! *J. Mol. Graphics. Model.* **2002**, *20*, 269−276.
(53) Buchbauer, G.; Spreitzer, H.; Zechmeister-Machhart, F.; Klinsky, A.; Weiss-Greiler, P.; Wolschann, P. Synthesis and olfactoric activity of keto-*β*-santalol and methoxy-*β*-santalol. *Eur. J. Med. Chem.* **1998**, *33*, 463−470.

(54) Buchbauer, G.; Hillisch, A.; Mraz, K.; Wolschann, P. Conformational parameters of the sandalwood-odor activity: conformational calculations on sandalwood odor. Part X. *Helv. Chim. Acta* **1994**, *77*, 2286−2296.
(55) Rossiter, K. Structure-Odor Relationships. *Chem. Rev.* **1996**, *96*, 3201−3240.
(56) Chastrette, M. Trends in structure-odor relationships. *SAR QSAR Environ. Res.* **1997**, *3*−*4*, 215−254.
(57) Zakarya, D.; Cherquaoui, D.; Essefar, M.; Villemin, D.; Cense, J. M. Application of neural networks to structure-sandalwood odor relationships. *J. Phys. Org. Chem.* **1997**, *10*, 612−622.
(58) Buchbauer, G.; Sunara, A.; Weiss-Greiler, P.; Wolschann P. Synthesis and olfactoric activity of side-chain modified β-santalol analogues. *Eur. J. Med. Chem.* **2001**, *36*, 673−683.
(59) Chapuis, Ch.; Gautier, A.; Blanc, P.-A. Use of optically active isomers of (e)-3,3-dimethyl-5-(2,2,3-trimethyl-3-cyclopenten-1-yl-)-4-penten-2-ol in perfumery. *EP 643958 A2*, **1995**, CAN 123:17516.
(60) Bajgrowicz, J. A.; Frater, G**.** Preparation of optically pure isomers of campholenic aldehyde derivatives for use as detergent fragrances. *EP 841318 A2*, **1998**, CAN 129:28103.
(61) Naipawer, R. E. Cyclopentene derivatives and their use as odorants. *EP 203528 A2*, **1986**, CAN 106:175828.
(62) Naipawer, R. E.; Easter, W. M. 3-Methyl-5-(2,2,3-trimethylcyclopent-3-en-1-yl)pentan-2-ol compound and perfume compositions. *US 4052341*, **1977**, CAN 88:22229.

CI020296N