

# A Rapid Method for Molecular Shape Comparison of Medium-Sized Molecules. Conformational Calculations on Sandalwood Odor, IV [1]\*\*

Armin Becker<sup>1</sup>, Gerhard Buchbauer<sup>1</sup>, Susanne Winiwarter<sup>1</sup>, and Peter Wolschann<sup>2,\*</sup>

<sup>1</sup> Institut für Pharmazeutische Chemie, Universität Wien, A-1090 Wien, Austria

<sup>2</sup> Institut für Theoretische Chemie und Strahlenchemie, Universität Wien, A-1090 Wien, Austria

**Summary.** A fast method of a surface comparison of two or more molecules to be matched is presented. The Van der Waals surfaces of molecules are described by points calculated as the intersection of grid lines with the molecular surface. The mean surface of various molecules with the same biological activity can be constructed. It is used for further comparisons with similar molecules lacking this activity. Deviations of any molecular surface from the mean surface can be mapped onto the surface. The method was tested on a distinct group of sandalwood odor molecules and it was shown that such matching and comparison procedures are useful in the investigation of odor structure-activity relationships proposed as CAFD (computer aided fragrance design).

**Keywords.** Molecular shape; Molecular surface comparison; Van der Waals surface; Computer aided fragrance design; Sandalwood odor.

**Eine schnelle Methode für den Vergleich von Molekülformen von mittelgroßen Molekülen. Konformationsanalytische Berechnungen an Sandelholzriechstoffen, 4.Mitt.**

**Zusammenfassung.** Es wird eine schnelle Methode zum Oberflächenvergleich von zwei oder mehr aneinandertzupassenden Molekülen vorgestellt. Die Van der Waals-Oberflächen werden als Kreuzungspunkte von Netzlينien mit der Moleküloberfläche beschrieben. Es können die mittleren Oberflächen verschiedener Moleküle mit der gleichen biologischen Aktivität konstruiert werden. Das wird für weitere Vergleiche mit ähnlichen Molekülen benutzt, denen die entsprechende Aktivität fehlt. Abweichungen von einer beliebigen molekularen Oberfläche von der mittleren Fläche kann auf dieser graphisch dargestellt werden. Die Methode wurde an einer bestimmten Gruppe von Molekülen mit Sandelholzgeruch getestet. Es wurde gezeigt, daß die Einpassungs- und Vergleichsprozedur für die Untersuchungen von Struktur-Aktivitäts-Beziehungen vom CAFD-Typ nützlich ist (CAFD = Computer Aided Fragrance Design).

## Introduction

Methods for comparing the molecular shapes of two or more molecules have received much attention in the investigation of structure-activity-relationships. The

---

\*\* Dedicated to Prof. Dr. W. Fleischhacker, on occasion of his 60th anniversary

search for common structural elements, which are responsible for a distinct biological effect, and the identification of surface regions, that might fit to a complementary receptor, led to the developments of different methods [2–18]. Overlapping volumes as well as accessible surfaces of two or more molecules were used in several cases as appropriate criteria for superimposition [8–17]. The comparison of two different molecular surfaces in 3D space is connected with some difficulties. The superimposition of molecules, necessary for a relevant comparison of their surfaces is a multidimensional minimization problem, possibly with many local minima [18]. The complete minimization procedure is therefore rather complex and needs a lot of computer time. In addition the quantitative description of molecular similarities or dissimilarities is rather sophisticated: surface comparisons need not be sufficient, in many cases other molecular properties like electrostatic potentials have to be considered, too. In the present study a rapid method of matching molecules and measuring the similarities between molecular surfaces of various different molecules held in a defined orientation is described. We applied this method to the investigation of the structure-activity-relationship of sandalwood odor molecules, in order to find the common structural element which might be responsible for the pronounced biological effect.

## Method

Any molecular comparison requires an exact conformational analysis of each of the molecules in question. We used an empirical force field program (MOLBMEC [19], based on Allinger's MM2 force field [20]) for calculating the different conformations and their energies. Additionally a semiempirical method (AM1 [21, 22]) was used for the evaluation of the electronic properties of the molecules, like dipole moments or electron density distributions.

All geometrical data of each conformation, i. e. the Cartesian coordinates of the atoms, a connectivity table and a list of attached atoms, were stored in a data base. The conformations were ordered according to their energies. Relevant conformations, which can be assumed to be appreciably populated at thermodynamic

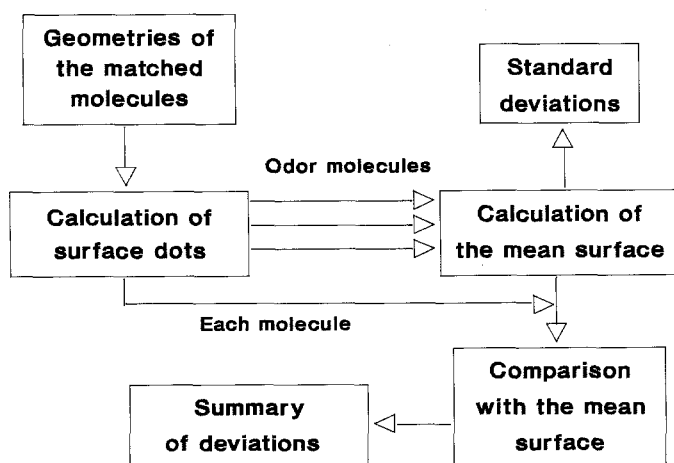


Fig. 1. Scheme of the calculation procedure (referring to the program packages used)

equilibrium, can be selected easily. Only the conformations with an energy not higher than 5 kcal/mol above the global minimum were taken into account.

As shown in Fig. 1, in the next step the molecules, respectively their different conformations, had to be oriented alike.

To avoid blind surface pattern matching, both molecules were compared in such a way that functional groups or hydrophobic or hydrophilic moieties which might be important to the reactivity of the molecules were matched first. In order to achieve this goal each molecular conformation had to be superimposed onto a "standard". This standard is represented by one definite conformation. It was selected for both its rigidity and its biological activity. The program GUDCON was developed especially for this task. Similar to other matching-routines [2, 3, 5–7], it fits only a few atoms of each molecule using a least squares algorithm, which gives the r. m. s.-value as the first criterion of the quality of the molecular matching. Of course, this value depends very much on the atoms one has selected for the superimposition. Therefore, GUDCON is able to choose the best fitting atoms out of a given set.

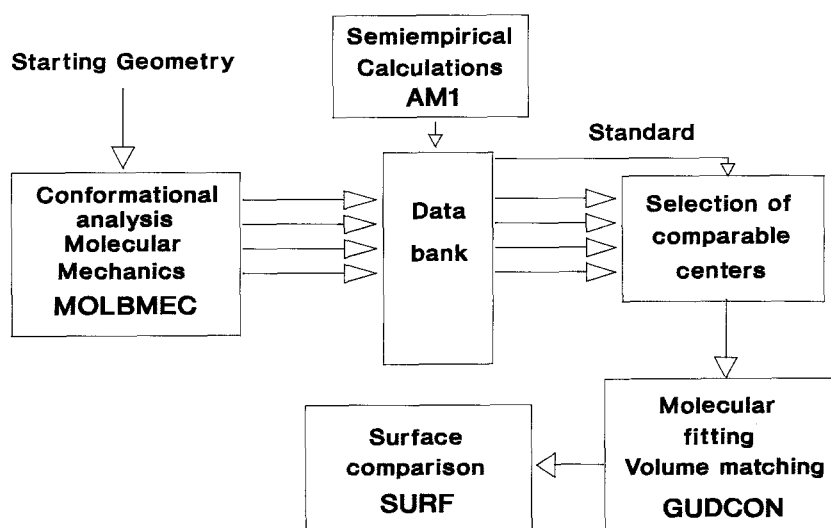


Fig. 2. Scheme of the surface comparisons

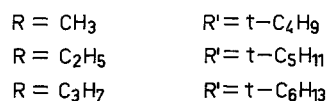
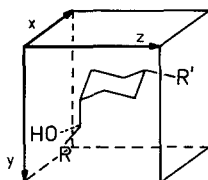


Fig. 3. Position of the molecules under investigation according to the coordinate axes

Another feature is the possibility to define maximum distances between certain atoms respectively of bonds between them. For instance, it can be useful to fix the orientation of a polar group, e. g. a hydroxy group in such a way that the carbon and oxygen atom of both molecules will be closer to each other than the other atoms that should be matched.

Still, when considering all relevant conformations, the further calculations would need quite a lot of computer time. So one needs a good quality criterion to eliminate conformations of minor interest. The r. m. s.-value in fact is sufficient only to serve as first criterion, but as it is based on only a few atoms, this is rather arbitrary and thus not really appropriate. With GUDCON one can also calculate and compare molecular volumes or even try to match two volumes. But it has to be noted that especially the latter is time-consuming and should be done only when the initial matching of the molecules is rather satisfying.

Only when reasonably good positional correspondence of the two molecules under investigation has been established, a surface comparison can be performed by means of the program module SURF. This program calculates the surface of a given conformation and compares also the surface pattern of two or more structures as shown in Fig. 2.

SURF calculates the surface of each molecule as a set of points which are obtained from the intersection of an exactly defined grid and the Van der Waals surface of the molecules. Thereby we used the Van der Waals radii proposed by Motoc [23]. For reasons of simplifications these lines are all defined parallel to one of the Cartesian axes as shown in Fig. 3.

The proper distance between grid lines was found to be about 0.2 Å. The corresponding surface is then represented by approximately 1 000 points, naturally depending on size and shape of the molecule. As each axis has actually two directions two pictures for representing the surface are necessary, when using lines parallel to one axis. So altogether six pictures for each molecule are obtained.

For comparison a mean surface of active substances is calculated at first. This computation is rather simple, as only the points of one line have to be considered when calculating the mean value for this line. In the same step also the mean deviations of the surface dots are computed and the values can be used for the colouring of the graphical representation of the mean surface points. So regions of the surfaces which are alike can be distinguished very easily from those that are dissimilar.

The next step is to compare this mean surface to the surfaces of active as well as of inactive compounds. The deviation of each point of the molecular surface of a certain conformation from the mean surface can be estimated, too. For a graphical representation similar to that one mentioned above the deviation of each point is mapped onto the surface. Regions of similarity of the various surfaces with the mean surface or patches of dissimilarities, with positive or negative deviations, can be differentiated by these calculations. A positive deviation may completely hinder an association of the molecule to a possible acceptor protein, a negative deviation diminishes the association constant and may therefore lead to loss of the biological activity, too.

A more quantitative measure of the differences between various molecular surfaces may be achieved by counting all points which are placed within a certain distance of the mean surface. These numbers are expressed in percentages and can be compared for all molecules.

### Calculations on Odor Molecules

The chemical mechanisms of odor perception and discrimination is not known up to now. The only information about this specific and extremely sensitive biological effect concerns some proteins inside the nasal epithelium [24–27]. It was found that the affinity of certain odor molecules to such receptor proteins is definitely higher than their affinity to other proteins [28]. The amino acid sequence as well as the structure of these receptor proteins is completely unknown. In order to obtain more insight into the physicochemical mechanism of odor recognition, a series of odor molecules with a typical fragrance, sandalwood odor, was studied [29–31]. The class of sandalwood fragrance consists of more than 70 compounds, showing this very sensitive and specific odor. Interestingly different substances with various functional groups and structure elements could be found. A comparison of the molecular shape of these molecules may lead to a common structural element which might be responsible for the biological effect.

For the present study one special group of sandalwood odor molecules was considered. All of them consist of one cyclohexane ring and two different residues, an aliphatic residue ( $R_1$ ) and one with a functional group like a hydroxyl or a carbonyl group ( $R_2$ ). The compounds differ only in these two residues and their mutual positions. A survey of the calculated odor compounds together with odorless compounds of similar structure is given in Fig. 4.

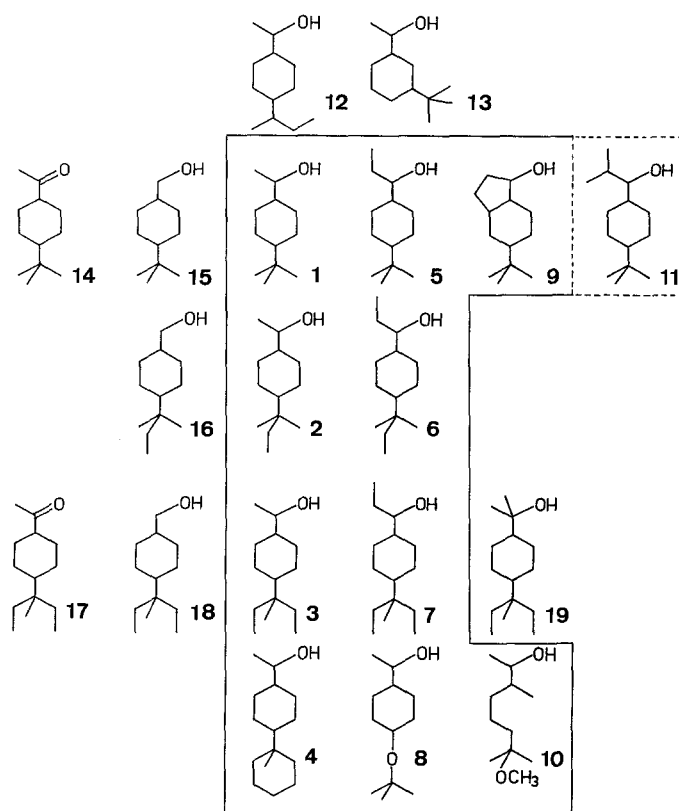


Fig. 4. Compounds under investigation (structures within the box show sandalwood fragrance)

Some structural requirements for odoriferous properties of this class of compounds have been postulated by Naipawer [32]:

1) One residue,  $R_1$ , has to be a tertiary alkyl-group of 4 to 6 carbon atoms (*t*-butyl, *t*-amyl or *t*-hexyl).

2) The second residue,  $R_2$ , must be either a 1-hydroxyethyl- or 1-hydroxypropylgroup (the former shows a significantly stronger sandalwood odor).

3) These two residues must be *para* positioned in the cyclohexane ring, their relative position has to be *cis*.

The compounds **1–3** and **5–7** follow these empirical rules (see Fig. 4), but even small changes in their structures can cause the total loss of the fragrance. E. g. in the case of primary alcohols, if  $R_1$  is only a hydroxymethyl-group (instead of a hydroxyethyl- or hydroxypropylgroup), no sandalwood odor can be recognized. Also a tertiary alcoholic group, like in **19**, cannot act as osmophoric group. Similarly all the other non-smelling substances in Fig. 4 just do not fit in the rules described above: In **20**, the *trans* isomer of **1**, the two residues are *trans* positioned instead of *cis*, compounds **14** and **17** bear an aldehyde function instead of a hydroxyl group and in **12** the alkyl residue is not a tertiary alkyl group but only a secondary one.

On the other hand, one finds similar substances which are not described by Naipawer's rules, but still show the typical sandalwood fragrance: osyrol<sup>R</sup> (**10**) does not consist of a cyclohexane ring, it is one of the few acyclic sandalwood odor molecules, **9** is a bicyclic derivate. In the ether **8** the tertiary alkyl group is connected to the cyclohexane ring by an intermediating oxygen and in **4** the tertiary alkyl group is replaced by an additional aliphatic ring, which still includes a quarternary carbon atom.

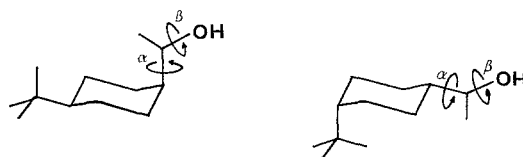
In the conformational analysis of these compounds, performed as described above, all possible rotations of single bonds as well as both different chair conformations of the cyclohexane ring were considered. When analysing **1**, *cis*-1-hydroxyethyl-4-*tert*-butyl-cyclohexane, we find 9 conformers for each cyclohexane chair conformation, so altogether 18. When analysing its *trans*-isomer, **20**, one of the chair conformations can be neglected, as the energy of a structure with two residues in axial position is *a priori* higher than one with both in equatorial position. But it was also found that the energy of all conformers of the *cis* compound with the hydroxy ethyl group in axial position is definitely lower than the others. The difference between the lower conformers of both groups is approximately 3–4 kcal/mol. Even if there is a hydroxypropyl group instead of the hydroxyethyl group (**5**) this difference can be found. Therefore also for all other *cis*-configured molecules only those with the hydroxyalkyl group in axial-position have been taken into account.

The next step was to match all the molecules in question with one standard; in the present study **1** was chosen as standard. This molecule is not completely rigid, but the rotation of the *tert*-butyl group leads to identical conformations and can be neglected. In addition a typical and strong sandalwood odor is described for this molecule [32]. The results of the conformational analysis of **1** are given in Table 1.

For this molecule 9 different conformers were taken into account, since only one chair conformation of the cyclohexane ring has sufficiently low energy. It can be shown that the rotation of the hydroxyl group has only neglectable influence

**Table 1.** Relative energies of the various conformations of **1** (angles in degrees, energies in kcal/mol)

| $\alpha$ | $\beta$ | energy | rel. ener. | $\alpha$ | $\beta$ | energy | rel. ener. |
|----------|---------|--------|------------|----------|---------|--------|------------|
| 61.0     | 99.0    | 19.76  | 1.83       | 62.4     | 66.1    | 22.28  | 4.36       |
| 60.6     | 168.1   | 19.43  | 1.50       | 64.4     | 168.1   | 22.31  | 4.39       |
| 58.6     | 271.8   | 20.27  | 2.34       | 67.3     | 285.4   | 22.87  | 4.95       |
| 202.0    | 70.5    | 21.43  | 3.50       | 180.2    | 85.1    | 21.86  | 3.94       |
| 203.2    | 166.2   | 21.39  | 3.46       | 180.9    | 166.5   | 21.71  | 3.80       |
| 208.4    | 293.2   | 21.81  | 3.88       | 182.3    | 298.4   | 22.12  | 4.20       |
| 302.4    | 84.90   | 18.02  | 0.09       | 292.4    | 77.0    | 22.00  | 4.08       |
| 302.6    | 167.0   | 17.93  | -          | 295.8    | 168.9   | 21.93  | 4.01       |
| 304.3    | 298.1   | 18.36  | 0.43       | 298.8    | 286.5   | 22.50  | 4.58       |



on the energy of the molecule. So there are just three different conformations, out of which **1** – **8**, the lowest energy conformation, was selected as standard.

All other molecules now were fitted as well as possible to this compound by means of GUDCON. As all of them belong to one group it was not too difficult to find atoms which should match: Evidently all the functional groups (in most cases a carbinol group) of each molecule should be oriented alike, so the oxygen and carbon atoms had to be matched to those of **1** – **8**. As third atom the quarternary carbon atom of the aliphatic residue (in the molecules as well as in the standard) was chosen. Only for **8** the oxygen atom of this residue was taken.

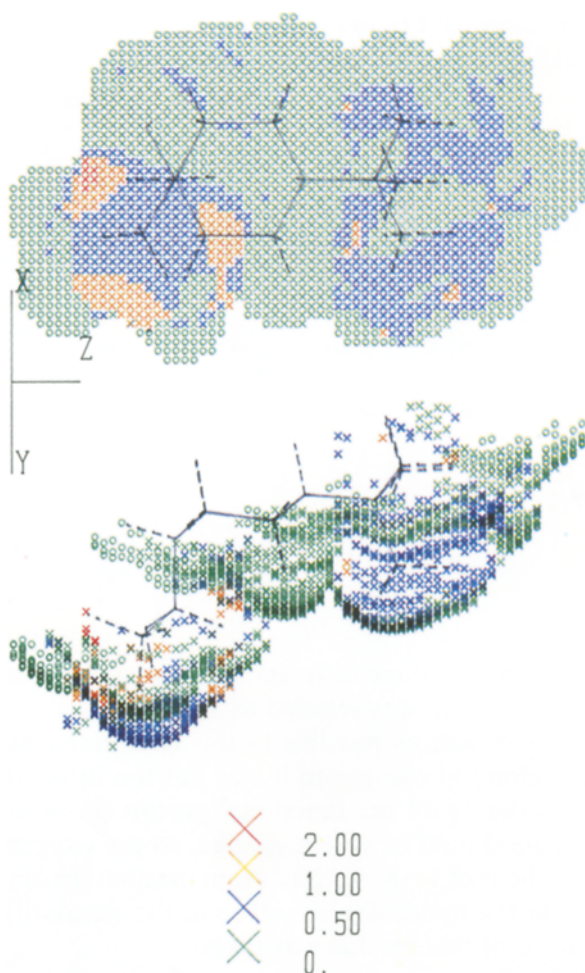
Before using the program SURF it was necessary to determine the best fitting conformation of each molecule. Several criteria were used for this selection:

First of all the graphical representation of the matched molecules allows to decide visually whether they fit well or not. As second criterion the r. m. s.-value of the calculation should be as low as possible, followed by a third criterion for the selection of a distinct conformation, namely the energy. If the first two criteria did not lead to one pronounced best fitting conformation the one with the lowest energy was selected. Only this conformation of each molecule was used for the further surface calculation.

For each direction the mean surface was calculated from the surfaces of the sandalwood odor molecules. At the same time a standard deviation was computed for every point. When represented graphically, 6 pictures were obtained according to the six different views in the program. One of them is shown in Fig. 5.

Points with large standard deviation are represented by red symbols whereas orange and blue ones indicate lower deviations. Green dots symbolize rather small deviations of all surfaces used for the estimation of the mean surface. So the regions where the surfaces of all molecules are similar can be distinguished easily from patches where the molecular surfaces differ rather strongly.

Afterwards the comparison of the surfaces of all compounds, also the odorless ones, with the mean surface was performed. As before a graphical representation

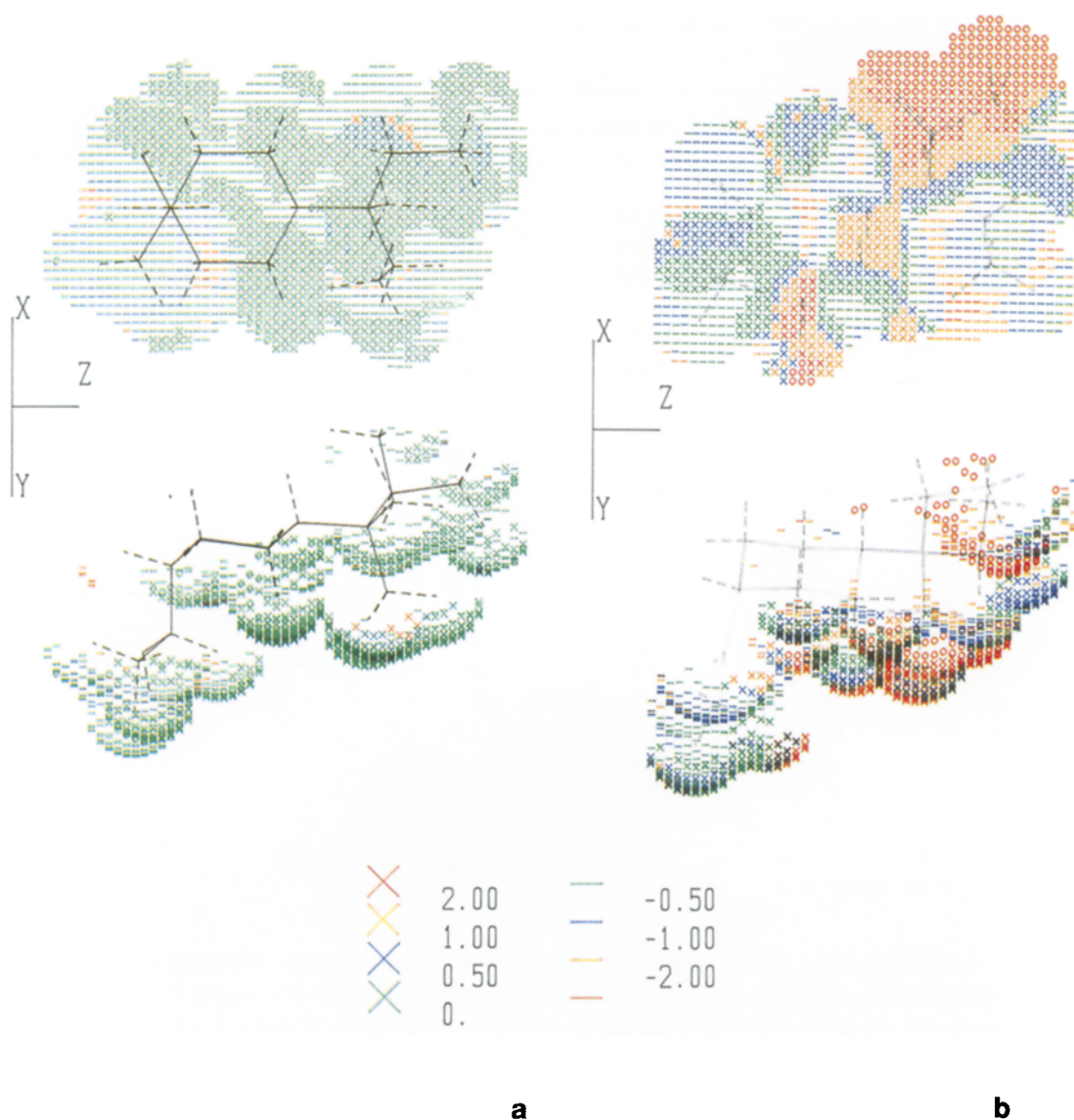


**Fig. 5.** Projection of the mean surface in positive  $y$ -direction, viewed in the  $y$ -axis and perpendicular to the  $y$ -axis. The standard deviations are mapped onto the surface

shows the result quite good. Here the deviation of each point from the corresponding point of the mean surface is mapped onto the surface. Green symbols indicate points very close to the mean surface, blue, orange and red ones those with a larger deviation in increasing order. A “+” shows a point which is “outside” the mean surface whereas a “-” symbolizes one within (an “o” can either indicate a point exactly on the mean surface - then it is green - or one not having a corresponding point on it - then it is red). Most odor compounds show a rather “green picture” in contrast to the odorless substances of which we mostly find “colourful” ones. As examples, the projections of an odorant molecule (**3**) and of a nonodorant molecule of similar structure (**17**) are shown in Fig. 6 a and 6 b. No large deviation can be observed in **3** whereas in **17** there is a large red region, caused by the different orientation of the aldehyde group in comparison to the position of the hydroxyl groups of the odoriferous molecules. The shape of the molecule **17** is therefore changed in such a way that an association to a receptor is drastically hindered.

The deviations from the mean surface were calculated for each compound shown





**Fig. 6. a** Projection of the surface of **3** in comparison to the mean surface (in positive  $y$ -direction). The deviations are mapped onto the surface (crosses indicate positive deviations, minus signs indicate negative deviations). The projections are viewed in the  $y$ -axis and perpendicular to the  $y$ -axis; **b** Projection of the surface of **17** in comparison to the mean surface

in Fig. 4, whether they were sandalwood odor molecules or not. In Table 2 the amount of surface dots with deviations less than  $1.5 \text{ \AA}$  is listed for each molecule, as well as the amount of dots that are farther than  $1.5 \text{ \AA}$  away from the mean surface, either inside or outside of it (given in %).

Negative deviations from the mean surface indicate that the given molecule is in distinct regions smaller than the average sandalwood odor molecule. This may cause a less stable complex with an appropriate receptor molecule, as no or di-

**Table 2.** Deviation of the individual surfaces from the mean surface in %

| Compound        | % deviation | % deviation   |        |
|-----------------|-------------|---------------|--------|
|                 | -1.5 Å      | -1.5 - +1.5 Å | +1.5 Å |
| 1               | 1.06        | 98.94         | 0.0    |
| 3               | 1.08        | 98.92         | 0.0    |
| 2               | 0.95        | 99.05         | 0      |
| 4               | 0.82        | 99.28         | 0      |
| 5               | 0.07        | 96.75         | 3.18   |
| 7               | 0.07        | 97.30         | 2.63   |
| 6               | 0           | 97.24         | 2.76   |
| 8               | 5.51        | 93.84         | 0.65   |
| 11              | 0.14        | 95.48         | 4.38   |
| 9               | 4.99        | 92.99         | 2.02   |
| 10              | 5.80        | 92.57         | 1.63   |
| 12              | 8.09        | 91.91         | 0      |
| 19              | 2.61        | 89.07         | 8.32   |
| 14              | 2.52        | 81.80         | 15.68  |
| 17              | 5.24        | 80.75         | 14.01  |
| 15              | 10.33       | 89.67         | 0      |
| 18              | 9.50        | 90.50         | 0      |
| 16              | 10.22       | 89.63         | 0.15   |
| 20 <sup>a</sup> | 12.62       | 78.47         | 8.91   |
| 21 <sup>b</sup> | 12.54       | 72.14         | 15.32  |

<sup>a</sup> *trans* isomer of **1**<sup>b</sup> *trans* isomer of **8**

minished contact with its surface is possible. Positive ones suggest parts on the molecular surface, which may hinder an association of the molecule to the receptor sterically. Sandalwood odor molecules (first part of Table 2) show a better agreement of their surfaces to the mean surface than other molecules of similar structure.

The surface parts of the compounds **1–11**, all sandalwood odor molecules, with a deviation of less than 1.5 Å are definitely larger (more than 92% of the total surface) than those of the other molecules which do not smell alike. Additionally the deviations of the surfaces of compounds **1–11** from the mean surface with values higher than 1.5 Å, either positive or negative, are significantly less (5.8% is maximum) than of the nonodorant molecules (at least 8.0%). E. g. compound **12**, which is in principle a substructure of **2**, shows definite bigger regions of negative surface deviations than **2**, which seems to cause the difference of the fragrance. Also the primary alcohols **15**, **16** and **18**, subunits of **1**, **2** and **3**, show patches of negative surface deviations, which fill about 10% of their total surface. Compounds **14**, **17**, **19**, **20** and **21** show rather big surface parts with positive deviations of more than 1.5 Å.

Still a definite boundary line between the two groups of molecules cannot be found, especially as the odor impression itself is very difficult to be defined exactly. E. g. the odor of **12** is described as green, warm and woody and resembling to sandalwood odor [32].

The presented fast method of molecular surface comparison has been applied and tested on a group of sandalwood odor molecules and it was shown that such a "coarse grained" method seems to be useful in the investigation of complex structure-activity (here fragrance) relationship.

### Acknowledgements

The authors are indebted to Prof. P. Schuster for helpful comments and the Hochschuljubiläumsstiftung der Stadt Wien for financial support. We want to thank the Rechenzentrum der Universität Wien for computer time on the IBM 3090. We also acknowledge gratefully the friendly interest in our work by DRAGOCO-Vienna.

### References

- [1] Becker A., Buchbauer G., Winiwarter S., Wolschann P. (1990) *J. Ess. Oil Res.* **2**: 221
- [2] Nyburg S. C. (*Acta Cryst.*) 1974 **B 30**: 251
- [3] Stouch T. R., Jurs P. C. (1986) *J. Chem. Inf. Comput. Sci.* **26**: 4
- [4] Rohrbauch R. H., Jurs P. C. (1987) *Analyt. Chim. Acta* **199**: 99
- [5] Lejeune J., Michel A., Vercauteren D. P. (1986) *J. Comp. Chem.* **7**: 739
- [6] Richards W. G., Hodgkin E. E. (1988) *J. Mol. Graphics* **24**: 1141
- [7] McLachlan A. D. (1982) *Acta Cryst. A* **6**: 871
- [8] Namasivayam S., Dean P. M. (1986) *J. Mol. Graphics* **4**: 46
- [9] Chau P. L., Dean P. M. (1987) *J. Mol. Graphics* **5**: 97
- [10] Dean P. M. (1987) *Molecular Foundations of Drug-Receptor Interaction*, Cambridge University Press, Cambridge
- [11] Dean P. M., Chau P. L. (1987) *J. Mol. Graphics* **5**: 152
- [12] Dean P. M., Callow P. (1987) *J. Mol. Graphics* **5**: 159
- [13] Dean P. M., Callow P., Chau P. L. (1988) *J. Mol. Graphics* **6**: 28
- [14] Bladon P. (1989) *J. Mol. Graphics* **7**: 130
- [15] Arteca G. A., Jammal V. B., Mezey P. (1988) *J. Comp. Chem.* **9**: 608
- [16] Arteca G. A., Jammal V. B., Mezey P., Yadav J. S., Hermsmeier M. A., Gund T. M. (1988) *J. Mol. Graphics* **6**: 45
- [17] Colloch N., Mornon J. P. (1987) *J. Mol. Graphics* **5**: 170
- [18] Radack G. M., Badler N. I. (1989) *Comp. Vis., Graph. Im. Proc.* **45**: 380
- [19] Beyer A., Wolschann P., Becker A., Buchbauer G. (1989) *J. Mol. Struct.* **196**: 371
- [20] Allinger N. L. (1977) *J. Am. Chem. Soc.* **99**: 8127
- [21] Dewar M. J. S., Zoebisch E. G., Healy E. F., Stewart J. P. (1985) *J. Am. Chem. Soc.* **107**: 3902
- [22] Dewar M. J. S., Dieter K. M. (1986) *J. Am. Chem. Soc.* **108**: 8075
- [23] Motoc I. (1983) *Z. Naturforsch.* **38 A**: 1342
- [24] Fesenko E. E., Novoselov V. I., Krapivinskaya L. D. (1979) *Biochim. Biophys. Acta* **587**: 424
- [25] Lancet D., Pace U. (1987) *TIBS* **12**: 63
- [26] Anholt R. R. H. (1987) *TIBS* **12**: 58
- [27] Ohloff G. (1986) *Experientia* **42**: 271
- [28] Pevsner J., Trifiletti R. R., Strittmatter S. M., Snyder S. H. (1985) *Proc. Natl. Acad. Sci. USA* **82**: 3050
- [29] Brunke E. J., Klein E. (1982) in: Theimer E. T. (ed.) *The Science of the Sense of Smell*. Academic Press, New York, p. 397
- [30] Brunke E. J., Klein E. (1981) in: Mookherjee B. D., Mussinan C. J. *Essential Oils*. Allured Publ. Corp., Wheaton, Illinois, p. 83
- [31] Naipawer R. E., Puryycki K. L., Shaffer G. W., Erickson R. E. (1981) in: Mookherjee B. D., Mussinan C. J. (eds.) *Essential Oil*. Allured Publ. Corp., Wheaton, Illinois, p. 105

- [32] Naipawer R. E. (1988) in: Lawrence B. M., Mooharjee B. D., Wills B. J. (eds.) *Flavors and Fragrances: World Perspective*. Proc. of the 10th Congr. of Ess. Oil, Frag. and Flavors. Elsevier, Amsterdam, p. 805

*Received August 23, 1990. Accepted September 5, 1990*