Threshold-Based Structure−Activity Relationships of Pyrazines with Bell-Pepper Flavor

Gerhard Buchbauer,*,† Christian Th. Klein,‡ Bettina Wailzer,‡ and Peter Wolschann‡

Institute of Theoretical Chemistry and Molecular Structural Biology, University of Vienna, Währinger Strasse 17, A-1090 Vienna, Austria, and Institute of Pharmaceutical Chemistry, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

Quantitative structure activity relationships (QSAR) and comparative molecular field analysis (CoMFA) are applied in order to explain the aroma of 46 bell-pepper aroma compounds. Biological activities log(1/*c*) values are used, where *c* stands for the detection threshold value of the aroma compound in water. Results of conventional QSAR and CoMFA are both satisfactory in statistical significance and predictive ability. We construct a qualitative model using the graphic features of CoMFA together with the results of "classical" QSAR analysis, which is performed by multiple linear regression. Finally, the human olfactory detection threshold values of excluded pyrazines are successfully predicted. This makes CoMFA and QSAR two important tools for designing new aroma compounds and in elucidating the mechanism of odor−receptor interaction.

Keywords: *Correlation analysis; multiple linear regression; molecular descriptors; molecular fields*

INTRODUCTION

Many classes of aromatic compounds have been shown to possess key to flavor properties of foods. Pyrazines, a class of nitrogen-containing heterocyclic compounds, are important flavor ingredients. Their general structure is represented in Table 1. R_1 is defined as the substituent with the heteroatom.

Pyrazines as aroma compounds were identified in such heated foods as bread (Mulders et al., 1973), different meats (Wassermann, 1972), baked potatoes (Buttery et al., 1973), and coffee (Bondarovich et al., 1967), where they are formed during the Maillard reaction from reducing sugars and amino acids (Ho, 1996), but they also occur in fresh vegetables such as tomatoes, asparagus, beans, spinach (Murray, Whitfield, 1975), and bell-peppers (Buttery et al., 1969), as products of the secondary metabolism.

The sensory properties of substituted pyrazines have a broad spectrum, ranging from green, earthy, nutty, roasted, to bell-pepper or woody. The perception of flavor is a consequence of the interaction between the flavor molecule and a specific receptor localized in the olfactory mucosa. On the basis of DNA analyses, it is assumed that there exist 100 to 1000 different receptors which can be grouped into subclasses. Compounds of distinct structure bind to various subclasses with different affinities (Hildebrand, Shephard, 1997).

The gap between the knowledge of the primary structure and the three-dimensional geometry of olfactory receptors is large: while the sequence of some receptors is already known, no detailed structural elucidation exists up to now. This makes molecular modeling methods to be an important tool for the study of the aroma compound−receptor interaction.

The relationship between bell-pepper odor and chemical structure has been explored by several authors. Pittet and Hruza (1974) suggest an unsaturated nitrogen-containing heterocycle as a general structure with an alkyl group at position 2 (preferable an isobutyl group) and a methoxy group in position 1 (Table 1). Substitutions in other positions are postulated to be unfavorable for bell-pepper odor. Parliment and Epstein (1973) point out the importance of the size of the alkyl substituent R_2 for bell-pepper aroma, suggesting that larger substituents (up to 6−9 carbon atoms) favor the aroma impression. Buttery et al. (1976) propose a more general structural pattern: R_2 should be an alkyl group, preferably an isobutyl group and R_1 a heteroatom with an alkyl group. More recently, Masuda and Mihara (1988) mention that besides the hydrophobic interaction stemming from the alkyl group R_2, hydrogen bonding between the nitrogen atoms of the pyrazine ring and the heteroatom of R_1 as donors on one hand, and acceptors from the receptor-pocket on the other hand, should be important for bell-pepper flavor. Another model is suggested by Rognon and Chastrette (1994). They propose a common pattern and interaction model for the bell-pepper aroma, in which on the one side a hydrogen bond to the nitrogen atom N_1 and on the other side dispersion interactions with three substructures belonging to R_2 are involved. A fundamental investigation of structure−aroma activity relationships of threshold values of alkylpyrazines with earthy odor impression was performed by Grosch et al. (1999).

In the present work, structure−flavor relationships of pyrazine-based molecules with bell-pepper aroma are analyzed by means of two different methods, multiple linear regression (MLR), and comparative molecular field analysis (CoMFA), to obtain predictive models for bell-pepper aroma, as well as a better understanding of the associative mechanism of these aroma compounds with the receptor site (Cramer et al., 1988).

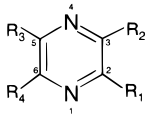
* To whom correspondence should be addressed. E-mail: Gehard.Buchbauer@univie.ac.at, Tel. +43 1 316368610.

† Institute of Pharmaceutical Chemistry.

‡ Institute of Theoretical Chemistry and Molecular Structural Biology.

10.1021/jf000192h CCC: $19.00
Published on Web 08/04/2000

Table 1. Structure of Pyrazines and Their Different Experimental log(1/*c*) Values



no.	R_1	R_2	R_3	R_4	log(1/*c*)	lit.
1	CH_3	C_3H_7	H	H	1.221	*f*
2	OCH_3	C_3H_7	H	H	3.920	*d*
					4.222	*i*
3	SCH_3	C_3H_7	H	H	3.000	*f*
4	CH_3	$CH(CH_3)_2$	H	H	1.795	*c*
5	OCH_3	$CH(CH_3)_2$	H	H	4.619	*d*
					5.698	*i*
					4.000	*a*
					6.000	*g*
6	OCH_3	$CH(CH_3)_2$	H	CH_3	4.346	*k*
7	OCH_3	$CH(CH_3)_2$	CH_3	H	4.301	*k*
8	OCH_3	$CH(CH_3)_2$	OCH_3	CH_3	1.154	*k*
9	OCH_3	$CH(CH_3)_2$	CH_3	OCH_3	0.602	*k*
10	OCH_3	$CH(CH_3)_2$	OCH_3	$CH(CH_3)_2$	0.173	*k*
11	SCH_3	$CH(CH_3)_2$	H	H	4.327	*f*
12	OCH_3	C_4H_9	H	H	4.301	*d*
13	SC_2H_5	C_4H_9	H	H	2.397	*d*
					1.215	*e*
14	CH_3	$CH_2CH(CH_3)_2$	H	H	0.886	*k*
15	OCH_3	$CH_2CH(CH_3)_2$	H	H	5.251	*j*
					5.698	*b*
					4.000	*a*
					4.348	*f*
					4.796	*k*
					3.745	*e*
16	OCH_3	$CH_2CH(CH_3)_2$	H	CH_3	3.585	*i*
17	OCH_3	$CH_2CH(CH_3)_2$	CH_3	H	2.585	*i*
18	OCH_3	$CH_2CH(CH_3)_2$	CH_3	CH_3	0.508	*i*
19	SCH_3	$CH_2CH(CH_3)_2$	H	H	3.481	*k*
20	OCH_3	$CH(CH_3)C_2H_5$	H	H	4.397	*d*
					5.699	*b*
					4.347	*h*
					6.000	*g*
21	OCH_3	C_5H_{11}	H	H	4.699	*d*
22	OC_2H_5	C_5H_{11}	H	H	4.096	*d*
23	SCH_3	C_5H_{11}	H	H	3.920	*d*
					5.000	*e*
24	SC_2H_5	C_5H_{11}	H	H	3.000	*d*
					2.602	*e*
25	OCH_3	$(CH_2)_2CH(CH_3)_2$	H	H	5.200	*h*
26	OCH_3	$CH_2CH(CH_3)C_2H_5$	H	H	4.920	*f*
27	OCH_3	$(CH_2)_3CH{=}CH_2$	H	H	4.522	*f*
28	OCH_3	$(CH_2)_2CH{=}CHCH_3$ (*E*)	H	H	3.886	*f*
29	OCH_3	$(CH_2)_2CH{=}CHCH_3$ (*Z*)	H	H	3.301	*d*
30	OCH_3	C_6H_{13}	H	H	4.154	*f*
					6.000	*j*
31	OCH_3	$(CH_2)_3CH(CH_3)_2$	H	H	5.221	*f*
32	OCH_3	$CH_2CH(CH_3)C_3H_7$	H	H	5.096	*d*
33	OCH_3	C_7H_{15}	H	H	4.585	*d*
34	OCH_3	C_8H_{17}	H	H	4.221	*d*
35	OC_2H_5	C_8H_{17}	H	H	2.699	*d*
36	SCH_3	C_8H_{17}	H	H	3.154	*d*
37	SC_2H_5	C_8H_{17}	H	H	2.699	*d*
38	OCH_3	$C_{10}H_{21}$	H	H	1.397	*d*
39	OC_2H_5	$C10H_{21}$	H	H	1.221	*d*
40	OCH_3	CH_3	OCH_3	CH_3	0.744	*g*
41	OCH_3	C_2H_5	H	H	high	*h*
42	OCH_3	$CH(CH_3)C_3H_7$	H	H	high	*h*
43	OCH_3	$(CH_2)_6CH(CH_3)_2$	H	H	low	*h*
44	OCH_3	$CH_2CH(CH_3)C_6H_{13}$	H	H	mod.	*h*
45	OCH_3	$CH_2CH(CH_3)_2$	H	$CH_2CH(CH_3)_2$	low	*h*
46	OC_2H_5	$CH_2CH(CH_3)_2$	H	H	mod.	*h*

[a] Calabretta (1978). [b] Buttery et al. (1969). [c] Flament; Stoll (1967). [d] Masuda, Mihara (1988). [e] Masuda, Mihara (1986). [f] Mihara, Masuda (1988). [g] Murray et al. (1970). [h] Parliment et al. (1973). [i] Seifert et al. (1972). [j] Seifert et al. (1970). [k] Takken et al. (1975).

EXPERIMENTAL PROCEDURES

Material and Methods. The chemical structures of 46 pyrazines with bell-pepper flavor are depicted in Table 1, together with their biological activities, expressed as log(1/*c*),

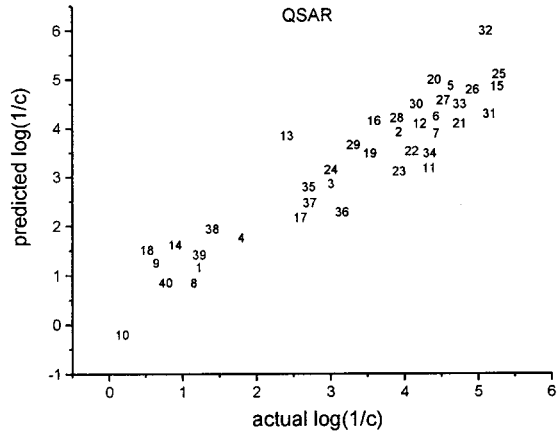


Figure 1. Experimental log(1/*c*) values plotted against predicted log(1/*c*) values obtained from 2D-QSAR analysis eq 1.

where *c* is the odor threshold value in water, given in ppm. In Table 1 threshold values of various authors are listed too. For the study the values from Masuda and Mihara (1988), Mihara and Misuda (1988) were taken mainly, because they were determined in the same way and are, therefore, much more comparable. The threshold values not used in the study are depicted in italic. The odor threshold means the olfactory detection threshold, that is the ability of the test person to distinguish between water with and without odorant. This should be distinguished from the recognition threshold, i.e., from the ability of a person to recognize the identity of the odor.

The 3D-structures of the compounds are built by the Hyperchem 5.0 software (Hypercube Inc., 1997) and fully minimized with the MM+ force field implemented therein. The resulting geometries are subsequently optimized with an ab initio Hartree−Fock method on 3-21G level, implemented in the GAUSSIAN 94 program (Gaussian, Inc., 1995).

Calculations of Properties for QSAR Analysis. For the structures obtained the following molecular properties were calculated, using the TSAR 3.21 (Tools for Structure−Activity Relationships) software (Oxford Molecular Limited, 1999).

1. Steric descriptors: molecular surface (*S*), molecular refractivity (MR), and the Verloop parameters (*L*, B_1, B_2, B_3, B_4, B_5) (Verloop, 1967) for the four substituents R_1 to R_4. (Figure 1). R_1 is always the substituent with the heteroatom, except for the three compounds, where it does not contain a heteroatom: compound **1**, **4**, and **14** (R_1 = methyl). $L^{(i)}$ represents the maximal length of substituent i along the axis defined by the bond, which connects the substituent with the heterocycle. $B_j^{(i)}$ ($j = 1,....,5$) denotes the widths of substituent i perpendicular to this axis and is chosen in such a way that $B_1 < B_2 < B_3 < B_4 < B_5$; the molecular refractivity, being related to the volume and the polarizability of a compound, is not only a steric descriptor, but also gives information on whether dispersion forces are important in the interaction with the receptor or not.

2. Descriptor of lipophilicity: log *P*, where *P* is the partition coefficient of a compound between octanol and water; the larger *P* (and thus log *P*), the more hydrophobic the compound. In the study, log *P* values of the different substituents (log P_1, ...log P_4) are used.

3. Electronic descriptors: Hartree−Fock-derived dipole moments (*D*) and point charges on the atoms of the heterocycle ($C^{(N1)}$, $C^{(N4)}$, $C^{(C2)}$, $C^{(C3)}$, $C^{(C5)}$, $C^{(C6)}$) and on the first atom of the respective substituent ($C^{(R1)}$, $C^{(R2)}$, $C^{(R3)}$, $C^{(R4)}$, see Table 1). The sum of electrotopological state indices, *E*, based on the electronegativity of an atom and its local topology, is another descriptor considered in the study.

Multiple linear regression (MLR) in TSAR is chosen to develop QSAR models, where the log(1/*c*) value is used as the dependent variable. The quality of the model is judged by the correlation coefficients (*r*), the standard error (*s*), the *t*-statistics of individual regression coefficient and the overall

Table 2. Correlation Matrix of the Significant Descriptors of Eq 1

	S	$C^{(R2)}$	$C^{(R3)}$	$L^{(R1)}$	$B_4^{(R1)}$	$B_5^{(R2)}$	log $P^{(R1)}$
S	1						
$C^{(R2)}$	−0.128	1					
$C^{(R3)}$	−0.097	−0.505	1				
$L^{(R1)}$	0.070	−0.220	0.041	1			
$B_4^{(R1)}$	0.415	−0.348	0.087	0.544	1		
$B_5^{(R2)}$	0.690	−0.266	0.190	−0.061	0.224	1	
log $P^{(R1)}$	0.084	−0.294	0.177	0.506	0.648	−0.057	1

F value. The reliability of the model is indicated, in terms of predictivity, by cross-validated r^2 (r^2_{cv}). Cross-validation is performed by leaving out each compound in turn (unless noted). MLR establishes a linear combination between the molecular properties of the molecules and their biological activities by determining the coefficients such, that the actual and predicted values are minimized.

CoMFA. CoMFA (Cramer 1988) is performed with the SYBYL 6.5 software (Tripos Inc., 1998). The molecules are superimposed by fitting the atoms of the heterocycle and the heteroatom. Grid sizes of 1.5, 2 and 2.5 Å and different probe atoms [$sp^3C(+1)$, $sp^3O(-1)$, $sp^3N(-1)$ and $H(+1)$] are employed for the evaluation of the molecular field. To reduce the amount of noise and to speed up partial least squares (PLS) analysis, the minimum σ value is set to 2.0 kcal/mol, and energy cutoff values of 30 kcal/mol are used for electrostatic and steric fields. For the calculation of the electrostatic field, the same atomic charges calculated by ab initio HF/3-21G as in MLR are used. PLS analysis is employed to correlate the field values with the biological activities. The quality of the models is estimated by the same statistical indicators as in MLR.

RESULTS

Multiple Linear Regression. For the multiple linear regression analysis, 40 aroma compounds with known threshold values are used (the first 40 from Table 1). The following regression model turned out to be the best:

$$\log(1/c) = -0.35S + 14.806C^{(R2)} + 2.503C^{(R3)} + 1.094L(R_1) + 1.443B_4^{(R1)} + 0.981B_5^{(R2)} - 3.174 \log P^{(R1)} + 4.186 \quad (1)$$

$$\log(1/c) = -1.293S^* + 0.742C^{(R2)*} + 0.795C^{(R3)*} + 0.421L^{(R1)*} + 0.712B_4^{(R1)*} + 1.288B_5^{(R2)*} - 0.830 \log P^{(R1)*} + 3.236 \quad (2)$$

$$s = 0.589, \quad r = 0.938, \quad F = 33.415, \quad r^2_{cv} = 0.811$$

where log(1/c) denotes the biological activity. The second equation (*-variables) is obtained by using data standardized to zero mean and unity variance, and thus the regression coefficients reflect the weights of individual contributions to the activity. In Table 2 the correlation matrix of the variables from eq 1 is listed. There is no correlation between the significant descriptors above 0.7.

A regression model can be considered to be significant if the overall F-value exceeds at least four times the percentage point $F_{\alpha, k, n-k-1}$ (Draper and Smith, 1981). For the above equation $F_{0.05, 7, 32} = 2.33$, i.e., the regression model is significant at the 95% level. Table 3 shows the statistics of the coefficients from eq 1. The t-values of the coefficients are all larger than the percentage point $t_{\alpha/2, n-k-1}$ ($t_{0.025, 32} = 2.042$), which shows, therefore, the significance of the individual regression coefficients

Table 3. Statistics of Significant Descriptors from Eq 1

descriptors	coefficient	t-value	standard error
S	−0.034	−8.544	0.004
$C^{(R2)}$	14.806	6.256	2.366
$C^{(R3)}$	2.502	6.781	0.369
$L^{(R1)}$	1.093	3.570	0.306
$B_4^{(R1)}$	1.442	4.653	0.310
$B_5^{(R2)}$	0.981	8.821	0.111
log $P^{(R1)}$	−3.174	−6.172	0.514

Table 4. Statistical Parameters and Significant Descriptors of Models Obtained from the Training Set (all 40 compounds) and Three Sets of 32 Compounds Each[a]

	training set all 40 compds)	test I	test II	test III
s	0.58	0.56	0.56	0.59
F	33.41	34.3	25.3	32.01
r^2	0.937	0.95	0.94	0.92
r^2_{CV}	0.811	0.84	0.74	0.72
$SDEP_e$		0.52	0.3	1.3
$SDEP_i$	0.53			
signif descript.	S, $C^{(R2)}$, $C^{(R3)}$, $L^{(R1)}$, $B_4^{(R1)}$, $B_5^{(R2)}$, log $P^{(R2)}$	S, $C^{(R2)}$ $C^{(R3)}$, $L^{(R1)}$, $B_4^{(R1)}$, $B_4^{(R2)}$, log $P^{(R1)}$	S, $C^{(R2)}$, $C^{(R3)}$, $L^{(R1)}$, $B_3^{(R1)}$, $B_4^{(R1)}$, $B_4^{(R2)}$, log $P^{(R1)}$	S, $C^{(R2)}$, $C^{(R3)}$, $L^{(R1)}$, $B_4^{(R2)}$, $B_5^{(R2)}$, log $P^{(R1)}$

[a] From the three test sets of eight compounds each, the external standard deviation of errors of prediction, $SDEP_E$, is calculated.

at the 95% level. The predictive power is high, since r_{cv}^2 is high and fairly close to r^2 (0.865).

The experimental versus the calculated affinities obtained from the QSAR model are plotted in Figure 1.

To analyze the external predictivity of the model, three test sets are generated by leaving out eight compounds at random from the initial data set of 40 compounds: 2, 3, 6, 11, 17, 20, 29, 35 (test set I); 4, 8, 16, 21, 26, 32, 37, 39 (test set II); 1, 2, 5, 9, 13, 18, 26, 34 (test set III). Each of the remaining three training sets of 32 compounds is used to produce QSAR models for the prediction of the activity of their respective test set of compounds. The external predictivity in terms of standard deviation of errors of prediction, $SDEP_e$, is in two of the cases lower than the internal standard deviation of errors of prediction, $SDEP_i$, of eq 1 (all 40 compounds). Moreover, the models yield more or less the same significant descriptors.

CoMFA Model. All analyses include both field types, i.e., steric and electrostatic fields. In Table 5 different models, which are obtained with CoMFA, are shown. We used grid sizes of 1.5, 2, and 2.5 Å and the following probe atoms: $sp^3C(+1)$, $sp^3O(-1)$, $sp^3N(-1)$ and $H(+1)$. The training set of 34 compounds is devised by excluding six compounds from the 2D-QSAR training set, namely 1, 7, 17, 23, 30, 38. The excluded compounds are used for external test predictivity.

CoMFA with default setting probe (sp^3C) and a grid size of 2 Å yields model 1 ($r_{cv}^2 = 0.65$). Using another probe atom, sp^3O (−1), produces a better r_{cv}^2 (model 6, $r_{cv}^2 = 0.63$), than those obtained from H (2 Å) and sp^3C (1.5 Å) (model 2, $r^2_{cv} = 0.52$, and model 5, $r_{cv}^2 = 0.53$). With sp^3N (−1) as probe atom the r_{cv}^2 is equal to 0.63 (model 3).

DISCUSSION

Results from the QSAR Analyses. The equation obtained from TSAR multiple regression analysis points out the importance of steric and electronic characteristics in contributing to the aroma intensity of the bell-pepper-odor pyrazines. The values of the regression

Table 5. Summary of CoMFA Models with Different Probe Atoms and Grid Sizes

model	grid size (Å)	probe atom	field type	r_{cv}^2	r^2	s	F	ster contr	Noc[b]
1	2	sp^3 C(+1)	both	0.65	0.98	0.21	254.6	70.5	7
2	2	H(+1)	both	0.52	0.97	0.28	160.0	64.8	6
3	2	sp^3 N(−1)	both	0.63	0.97	0.24	213.3	71.7	6
4	2.5	sp^3 C(+1)	both	0.64	0.98	0.19	285.8	74.5	7
5	1.5	sp^3 C(+1)	both	0.53	0.97	0.24	213.3	67.4	6
6	2	sp^3 O(−1)	both	0.63	0.97	0.24	214.7	71.5	6
1a	2	sp^3 C(+1)	both	0.733	0.977	0.25	185.43	65.5	6
1b	2	sp^3 C(+1)	both	0.782	0.996	0.11	757.7	67.5	8
1c	2	sp^3 C(+1)	both	0.809	0.994	0.14	525.5	68.0	7
1d	2	sp^3 C(+1)	both	0.841	0.997	0.09	1043.7	67.7	8

[a] 1a, elimination of cpd 39; 1b, elimination of cpd 6; 1c, elimination of cpd 35; 1d, elimination of cpd 2. [b] Number of components.

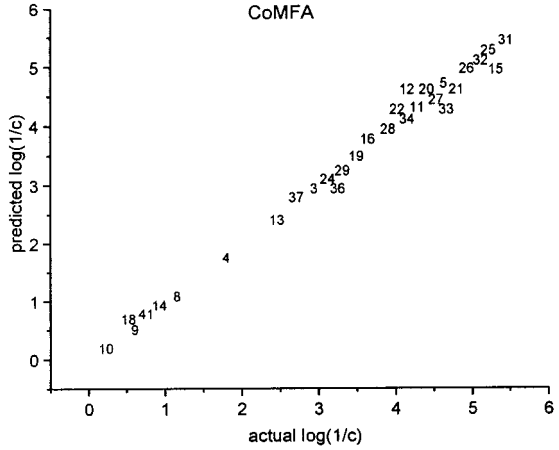


Figure 2. Plot of calculated versus experimental aroma intensity obtained from the CoMFA model 1d for the training set compounds.

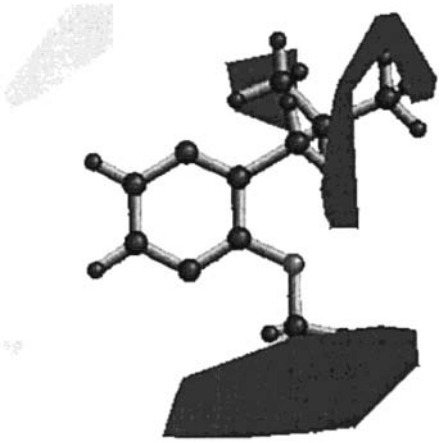


Figure 3. CoMFA steric contribution plot from the analysis based on the 3D-QSAR model 1d. Sterically favored areas are represented by dark-gray regions. Sterically unfavorable areas are represented by light-gray regions. Compound 15 is displayed inside the regions as a ball-and-stick.

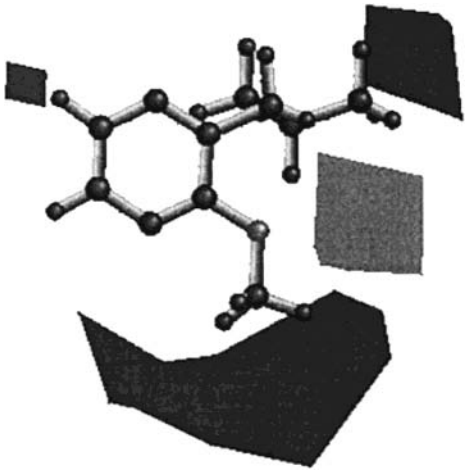


Figure 4. CoMFA electrostatic contribution plot from the analysis based on the 3D-QSAR model 1d. Negative charge favored areas are represented by light-gray regions. Negative charge unfavorable areas are represented by dark-gray regions. Compound 15 is displayed inside the fields as a ball-and-stick.

coefficients indicate the favorable steric contribution of all three descriptors ($L_1^{(R1)}$, $B_4^{(R1)}$, $B_5^{(R2)}$) to the biological activity. Bulky groups in the position of substituent R_2 and R_1 increase the biological activity, whereas the unfavorable contribution of bulky groups in the position of substituents R_3 and R_4, as modeled by CoMFA, is not reproduced by the MLR equation. Moreover, the equation suggests that increased positive charges at the first atoms of substituent R_3 ($C^{(R3)}$) are of advantage for bell-pepper flavor, because their values are mainly positive (except five of the compounds) and have positive regression coefficients (Table 3). This situation is more or less in agreement with the CoMFA picture (Figure 3). The unfavorable effect of a negative electrostatic field in the region of substituent R_2 results from the values of C^{R2} which are negative, and the regression coefficient which is positive, indicating that the more negative C^{R2} is, the lower the contribution to "bell-pepper flavor" will be. This is also in agreement with the CoMFA picture (Figure 4), where a positive field at R_2 favors the biological activity.

The molecular surface of the molecules should not be too high, because it leads to a decrease of the biological activity, which is described by a negative regression coefficient. The last significant descriptor turns out to be the log P value of substituent R_1. The favorable effect of low log P values at R_1 are suggested by the MLR regression analysis, because of the negative regression coefficient (Table 3).

Results of CoMFA. Model 1 yields the highest cross-validated r^2-value and the second lowest s-value (Table 5). Inspection of outliers shows that elimination of compound 39, 6, 35, and 2 could give a better prediction of the whole training set. Eliminating compound 2 from the CoMFA training set yields a model (1d − Table 5) with the highest cross-validated r^2 value of 0.841. The other statistical parameters are also very good: conventional r^2 is 0.997, the standard error is 0.09 and the overall F is 1044, indicating an excellent statistical significance of the model. Consequently, internal prediction (i.e., the prediction within the training set) is very good, as shown in Figure 2.

To predict the odor-threshold value of compounds in the test set, we used model 1d. In Table 6 the actual and calculated log(1/c) of nine test molecules are pre-

Table 6. Actual and Predicted log(1/*c*) Values of the Test Set by CoMFA

compound	actual log(1/*c*)	calculated log(1/*c*)	residual
1	1.222	0.799	0.423
7	4.301	2.364	−1.937
17	2.585	2.738	0.153
23	3.921	3.256	−0.665
30	4.155	4.594	0.439
38	1.398	4.556	3.158
41	high	4.379	
42	high	4.703	
45	low	0.496	

sented, where six of the nine structures have quantitative threshold values and three only a classification into high and low intensity. It can be observed, that compounds 1, 17, 23, 30, 41, 42, and 45 are quite well predicted. Completely wrong is the prediction of pyrazine 38, with a residual of 3.158. The reason for this is that there is no similar compound in the training set, no. 39 being excluded as an outlier in model 1d. The most similar is no. 34 with a log(1/*c*) of 4.22, which is close to the predicted 4.556 of no. 38. We obtained an exact prediction of compound 38 (residual = 0.292), when we included 39 in the training set, the model then having an r_{CV}^2 of 0.732.

The second pyrazine, which is not correctly calculated, is compound 7 with a residual of −1.937. Like compounds 9, 17, and 18, no. 7 has a methyl group in position R_4. The experimental log(1/*c*) values of these three compounds are lower than that of no. 7, and only no. 9 and 18 belong to the training set, which leads to a lower prediction.

Steric and Electrostatic Contributions. The 3D model for the bell-pepper aroma impression of the pyrazine derivatives in the training set is proposed in Figure 3. The model consists of steric (Figure 3) and electrostatic regions (Figure 4). Compound 15 is displayed inside the fields as a ball-and-stick model. The dark or light gray regions represent regions of space whose occupancy by ligand increases or decreases, respectively, the binding affinity for the receptor. There are favorable steric regions corresponding to the location of R_1 and R_2, which is shown in Figure 3. Introduction of bulky groups in these dark gray regions would increase the odor strength (lower human olfactory threshold value). On the contrary, bulkiness is unfavorable near the light gray contour field (in the position R_3 and R_4) for biological activity, i.e., for bell-pepper flavor.

The electrostatic contribution contour field is depicted in Figure 4. The positive electrostatic regions are shown in dark gray, and the negative contours are shown in light gray.

Positive charge in the region of the alkyl group of substituent R_1 and near the carbon atoms C_3−C_4 of substituent R_2 increases the biological activity. Positive charge should also be located in the region of R_3. A negative field is favorable in the vicinity of the heteroatom, i.e., a more negative heteroatom favors bell-pepper aroma.

The structure−odor relationship of pyrazine-based aroma compounds with bell-pepper flavor is successfully explained by 2D-QSAR analysis and CoMFA. The models obtained give information about the steric and electrostatic requirements for pyrazines to exhibit bell-pepper flavor. The results are based both on statistical significance and predictive ability, as reflected by the cross-validation r^2. The knowledge obtained is also in good agreement with the models already published.

ACKNOWLEDGMENT

We are grateful to DRAGOCO, Vienna, for their kind interest.

LITERATURE CITED

Buttery, R. G.; Seifert, R. M.; Guadagni, D. G.; Ling, L. C. Characterization of Some Volatile Constituents of Bell-peppers. *J. Agric. Food Chem*. **1969**, *17*, 1322−1327.

Buttery, R. G.; Guadagni, D. G.; Ling, L. C. Volatile Components of Baked Potatoes. *J. Sci. Food Agric*. **1973**, *24*, 1125−1131.

Buttery, R. G.; Guadagni, D. E.; Lundin, R. E. 4,5-Dialkylthiazoles with potent bell-pepperlike aromas. *J. Agric. Food Chem*. **1976**, *24*, 1−3.

Bondarovich, H. A.; Friedel, P.; Krampl, V.; Renner, J. A.; Shephard, F. W.; Gianturco, M. A. Volatile constituents of coffee. Pyrazines and other compounds. *J. Agric. Food Chem*. **1967**, *15*, 1093−1099.

Calabretta, P. J. Synthesis of Some Substituted Pyrazines and Their Olfactory Properties. *Perfum. Flavor*. **1978**, *3* (3), 33−42.

Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc*. **1988**, *110*, 5959−5971.

Draper, N. R.; Smith, H. The General Regression Situation. In *Applied Regression Analysis*; J. Wiley & Sons: New York, 1981; p 709.

Flament, I..; Stoll, M. Pyrazines. 1. Synthesis of 3-Alkyl-2-methylpyrazines by Condensation of Ethylenediamine with 2,3-Dioxoalkanes. *Helv. Chim. Acta* **1967**, *50*, 1754−1758.

GAUSSIAN 94, Revision B.3, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Rabb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavarachi, K.: Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, Foresman, J. B. M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andreas, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzales, C.; Pople, J.-A.; Gaussian, Inc.: Pittsburgh, PA, 1995.

Grosch, W.; Bielohradsky, J.; Czerny, M.; Wagner, R. Structure-odour-activity relationships of alkylpyrazines. *Z. Lebensm. Unters. Forsch. A* **1999**, *208*, 308−316.

Hildebrand, J. G.; Sheppard, G. G. Mechanisms of Olfactory Discrimination converging Evidence for Common Principles Across Phyla. *Am. Rev. Neurosci*. **1997**, *20*, 595−631.

Hall, L.; Mohney, B.; Kier, L. B. The electropological state: structure information at the atomic level for molecular graphs. *J. Chem. Inf. Comput. Sci*. **1991**, *31*, 76−82.

Ho, C. T. Thermal Generation of Maillard Aromas. In *The Maillard Reaction, Consequences for the Chemical and Life Science*; Ikan, R., Ed.; J. Wiley & Sons: Chichester, New York, Brisbane, Toronto, Singapore, 1996; pp 27−53.

Hyperchem 5.0, Hypercube Inc., 1997.

Masuda, H.; Mihara, S. Olfactive Properties of Alkylpyrazines and 3-Substituted 2-Alkylpyrazines. *J. Agric. Food Chem*. **1988**, *36*, 584−587.

Masuda, H.; Mihara, S. Synthesis of Alkoxy-, (Alkylthio)-, Phenoxy-, and (Phenylthio)pyrazines and Their Olfactive Properties. *J. Agric. Food Chem*. **1986**, *34*, 377−381.

Mihara, S.; Masuda, H. Structure−Odor Relationships for Disubstituted Pyrazines. *J. Agric. Food Chem*. **1988**, *36*, 1242−1247.

Mulders, E. J.; Ten Noever De Brouwn, M. C.; Van Straiten, S. The Odour of White Bread. *J. Food Res. Technol*. **1973**, *150*, 306−310.

Murray, K. E.; Whitfield, F. The Occurrence of 3-Alkyl-2-methoxypyrazines in Raw Vegetable. *J. Sci. Food Agric.* **1975**, *26*, 937−986.

Murray, K. E.; Shipton, J.; Whitfield, F. B. 2-Methoxypyrazines and the flavour of green peas. *Chem. Ind.* **1970**, *4*, 897−898.

Parliament, T. H.; Epstein, M. F. Organoleptic Properties of some alkyl-substituted alkoxy- and alkylthiopyrazines. *J. Agric. Food Chem.* **1973**, *21*, 714−716.

Pittet, A. O.; Hruza, D. E. Comparative Study of Flavor Properties of Thiazole Derivatives. *J. Agric. Food Chem.* **1974**, *22*, 264−269.

Rognon, C.; Chastrette, M. Structure-odor relationships: a highly predictive tridimensional interaction model for the bell-pepper note. *Eur. J. Med. Chem.* **1994**, *29*, 595−609.

Seifert, R. M.; Buttery, R. G.; Guadagni, D. G.; Black, D. R.; Harris, J. G. Synthesis and Odor Properties of Some Additional Compounds Relates to 2-Isobutyl-3-methoxypyrazine. *J. Agric. Food Chem.* **1972,** *20*, 135−137.

Seifert, R. M.; Buttery, R. G.; Guadagni, D. G.; Black, D. R.; Harris; J. G. Synthesis of some 2-Methoxy-3-Alkylpyrazines with Strong Bell-pepper-Like Odors. *J. Agric. Food Chem.* **1970**, *18*, 246−249.

SYBYL 6.5, Molecular Modeling Software, Tripos Inc., 1998.

Takken, H. J.; Van der Linde, M. L.; Boelens, M.; Van Dort, J. M. Olfactive Properties of a Number of Polysubstituted Pyrazines. *J. Agric. Food Chem.* **1975**, *23*, 638−642.

TSAR 3.2, Oxford Molecular Limited, 1999.

Wassermann, A. E. Thermally Produced Flavor Components in the Aroma of Meat and Poultry. *J. Agric. Food Chem.* **1972**, *20*, 737−741.

Received for review February 11, 2000. Revised manuscript received June 2, 2000. Accepted June 5, 2000. The authors thank the Fond zur Foerderung der wissenschaftlichen Forschung (project 11349-CHE) for financial support.

JF000192H