J Mol Model (2006) 12: 985–989
DOI 10.1007/s00894-006-0115-1

ORIGINAL PAPER

Driss Zakarya · Lahbib Farhaoui · Mohammed Hamidi · Mohammed Bouachrine

Structure–olfactive threshold relationships for pyrazine derivatives

Received: 29 October 2005 / Accepted: 14 March 2006 / Published online: 5 May 2006

Abstract Structure–olfactive threshold relationships for 40 pyrazine derivatives have been studied by multivariate statistical analysis. Variable descriptors used to describe the molecules studied were obtained using molecular-mechanics software. A correlation coefficient of 0.82 was obtained when all the molecules were included in the model. When the model was established for closely related subsets of molecules, the correlation coefficients obtained were higher and the established models were bilinear. Such models allow the identification of optimal structures corresponding to low olfactive thresholds for the subsets studied. Surprisingly, we find that the optimum structures are included in the set of 40 molecules. The efficiency of the models was supported by the cross-validation technique, where the correlation coefficients were found to be good with respect to the precision of the values of the olfactive thresholds.

Keywords Pyrazines · Olfactive thresholds · Multiple linear regression (MLR)

Introduction

Pyrazines are compounds that are well known to have remarkably low olfactive thresholds and interesting fragrances. Such compounds are usually found in cocoa [1] and human wastes. We were interested in studying pyrazines to determine how their olfactive activity depends on their chemical structure. If such a relationship is identified, we can understand which chemical groups are responsible for the biological behavior of the molecules.

Generally, experimental measurements of olfactive thresholds, like those of biological properties or behavior, are too expensive and require considerable time and expertise to be performed. To avoid such drawbacks, we try to establish mathematical models to obtain significant predictions of the properties or the biological activities of the compounds.

In a previous paper published by Zakarya [2], it was shown that olfactive thresholds are difficult to measure with sufficient precision. However, the model that Zakarya obtained was statistically significant for a set of alcohols. In this paper, we show that the olfactive threshold is related to physicochemical parameters used as molecular descriptors of the pyrazine chemical structures under study. The descriptors generated represent substructures attached to the common pyrazine skeleton. This approach is in agreement with the Hansch concept, which considers that, in many situations, variation of the chemical structure can lead to modification of the biological activity.

Materials and methods

Sample of pyrazines studied

The sample of compounds studied includes 40 pyrazine derivatives, given in Table 1. Their olfactive thresholds are taken from two main references [3, 4]. Chemical structures belong to the general structure (R_1)–CP–(R_2), where CP is a pyrazine ring, R_1 is a hydrocarbon chain and R_2 is OCH_3, OC_2H_5, SCH_3, SC_2H_5, or H (Fig. 1). R_1 and R_2 are located in positions 2 and 3 of the ring.

Descriptors used

In the field of chemometrics, many possibilities are offered to choose molecular descriptors. In our study, we chose the following descriptors according to the Hansh concept. As

D. Zakarya (✉) · L. Farhaoui
Department of Chemistry, Faculty of Sciences and Technology,
UFR of Applied Chemistry,
B.P 146, 20450 Mohammadia, Morocco
e-mail: zakaryadriss@yahoo.fr

M. Hamidi · M. Bouachrine
Faculty of Science and Technology,
Errachidia, Morocco

Table 1 Structure of radicals attached to pyrazine, olfactive thresholds of the molecules, and relevant parameters

N	R_1	R_2	Log(1/*t*)	d_1	d_2	*DM*	N_{1c}	N_{2c}	$V(R_1)$	*W*
1	H	H	3.523	1.331	2.36	0	−0.614	−0.614	0	1
2	CH_3	H	4.523	2.44	3.751	0.555	−0.6168	−0.6522	13.67	4
3	C_2H_5	H	5.398	2.782	4.913	0.53	−0.6181	−0.6685	23.9	10
4	C_3H_7	H	6.523	4.327	6.251	0.551	−0.6185	−0.6705	34.13	20
5	C_4H_9	H	6.398	5.197	7.471	0.566	−0.6186	−0.6709	44.36	35
6	C_5H_{11}	H	8.301	6.677	8.795	0.573	−0.6185	−0.6711	54.59	56
7	C_6H_{13}	H	6.699	7.708	10.033	0.58	−0.6187	−0.6711	64.82	84
8	C_7H_{15}	H	7.000	10.079	10.633	0.830	−0.6215	−0.6624	75.05	120
9	C_8H_{17}	H	6.398	10.245	11.705	0.857	−0.6215	−0.6624	85.28	165
10	$C_{10}H_{21}$	H	5.955	13.95	14.24	0.868	−0.6215	−0.6624	105.74	286
11	H	SC_2H_5	6.046	1.321	1.390	2.2033	−0.6073	−0.6483	0	20
12	CH_3	SC_2H_5	7.155	2.434	3.745	1.5755	−0.6461	−0.6533	13.67	35
13	C_2H_5	SC_2H_5	7.222	2.774	4.903	1.5818	−0.6547	−0.6632	23.9	56
14	C_4H_9	SC_2H_5	8.398	5.191	7.46	1.5558	−0.6554	−0.6657	44.36	120
15	C_5H_{11}	SC_2H_5	9.000	6.669	8.786	1.5638	−0.6555	−0.6660	54.59	165
16	C_8H_{17}	SC_2H_5	8.699	10.241	12.585	1.5664	−0.6556	−0.6661	85.28	364
17	$C_{10}H_{21}$	SC_2H_5	6.921	12.788	15.148	1.5565	−0.6555	−0.6661	105.74	560
18	H	SCH_3	6.699	1.321	2.354	2.089	−0.6068	−0.6459	0	10
19	CH_3	SCH_3	8.398	2.434	3.745	1.456	−0.6456	−0.6507	13.67	20
20	C_2H_5	SCH_3	7.398	2.774	4.903	1.481	−0.6628	−0.6521	23.9	35
21	C_3H_7	SCH_3	9.000	4.318	6.242	1.479	−0.6648	−0.6527	34.13	56
22	C_5H_{11}	SCH_3	9.921	6.67	8.786	1.473	−0.6655	−0.6528	54.59	120
23	C_8H_{17}	SCH_3	9.155	10.241	12.585	1.453	−0.6655	−0.6528	85.28	286
24	$C_{10}H_{21}$	SCH_3	7.699	12.789	15.148	1.454	−0.6655	−0.6529	105.74	455
25	H	OC_2H_5	7.097	1.313	2.358	1.936	−0.6081	−0.7269	0	10
26	CH_3	OC_2H_5	9.097	2.448	3.726	1.6253	−0.6537	−0.7309	13.67	35
27	C_2H_5	OC_2H_5	7.699	2.804	4.905	1.6099	−0.671	−0.7339	23.9	56
28	C_5H_{11}	OC_2H_5	10.097	6.703	8.779	1.584	−0.674	−0.7339	54.59	165
29	C_8H_{17}	OC_2H_5	8.699	10.277	12.587	1.6092	−0.6741	−0.734	85.28	364
30	$C_{10}H_{21}$	OC_2H_5	7.222	12.825	15.15	1.6129	−0.6706	−0.734	105.74	560
31	H	OCH_3	6.398	1.313	2.358	2.0071	−0.6079	−0.726	0	10
32	CH_3	OCH_3	8.155	2.382	3.8	1.647	−0.6535	−0.7299	13.67	20
33	C_2H_5	OCH_3	8.000	2.804	4.905	1.683	−0.6708	−0.7324	23.9	35
34	C_3H_7	OCH_3	9.921	4.34	6.246	1.617	−0.6732	−0.7327	34.14	56
35	C_4H_9	OCH_3	10.301	5.235	7.483	1.63	−0.6736	−0.7328	44.36	84
36	C_5H_{11}	OCH_3	10.699	6.703	8.779	1.62	−0.6738	−0.7329	54.59	120
37	C_6H_{13}	OCH_3	10.155	7.743	10.038	1.637	−0.6738	−0.7326	64.82	165
38	C_7H_{15}	OCH_3	10.585	9.17	11.333	1.627	−0.6738	−0.7328	75.05	220
39	C_8H_{17}	OCH_3	10.222	10.277	12.587	1.6428	−0.6738	−0.7328	85.28	286
40	$C_{10}H_{21}$	OCH_3	7.398	12.825	15.15	1.6428	−0.6739	−0.7329	105.74	455

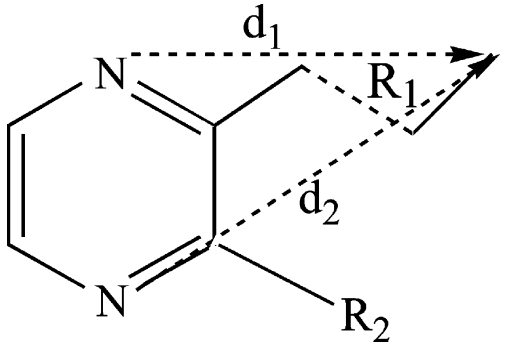

Fig. 1 General chemical structure of the molecules studied

all the molecules have a common substructure, the descriptors calculated are properties of R_1 and R_2 only. These properties are given below:

1. Contribution to the van der Waals volume [5] of R_1 represented by $V(R_1)$
2. Wiener index (*W*) of the R_1–R_2 branch [6].

These descriptors account for the size and branching of the chemical structure. They also account for the lipophilicity of the molecule because the chemical structures are closely similar.

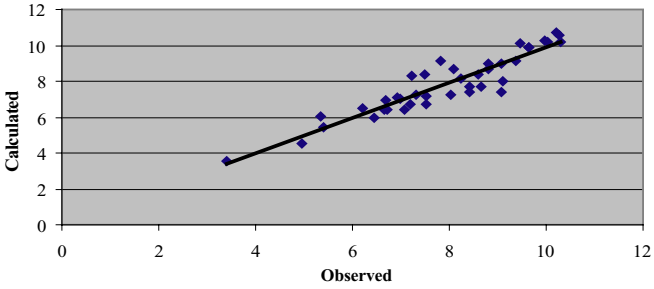

Fig. 2 Predicted vs observed Log(1/*t*)

1. Electrostatic charges located on nitrogen (*N*), and their real distance d_1 and d_2 to R_1 Group (Fig. 1). The charges on the nitrogen near R_1 and R_2 are represented by N_{1c} and N_{2c}, respectively.
2. Dipole moment of the molecule (*DM*)

The electrostatic and geometrical descriptors were determined by ab initio HF/3-21G calculations using the 98-Gaussian program.

Statistical analysis

The data obtained were submitted to statistical analysis to extract pertinent models. The method of statistical analysis used was regression analysis [2]. The pertinent variables of each model given below are statistically significant at 95%.

Results and discussions

As a first step, we tried to take into account all the molecules of the whole data set. The reason is to test the possibility of a general and unique model. The best equation obtained is given below:

$$Log(1/t) = -22.215 + 2.015d_2 + 1.999DM - 33.998N_{1c} - 0.176V(R_1) - 0.015W \quad (1)$$

N=40, r=0.93, and s=0.66; n is the number of molecules (sample size), r is the correlation coefficient, s is the standard error, and Log(1/*t*) is the olfactive threshold.

The model obtained Eq. (1) is interesting for the prediction of the olfactive threshold because it is statistically significant and includes molecular descriptors that are easy to compute. Indeed, W can be calculated from the molecular graph, and $V(R_1)$ is estimated from atomic contributions. Only the electrostatic and geometrical parameters (N_{1c}, *DM*, and d_2) need molecular mechanics

Table 2 Correlation coefficients between variables

	d_2	*DM*	N_{1c}	$V(R_1)$	*W*
d_2	1	0.063	0.43	0.99	0.9
DM		1	0.44	0.007	0.17
N_{1c}			1	0.41	0.46
$V(R_1)$				1	0.89
W					1

Table 3 Correlation coefficients between variables

	DM	N_{1c}	N_{2c}	*W*
DM	1	0.25	0.29	0.01
N_{1c}		1	−0.38	0.42
N_{2c}			1	−0.08
W				1

Table 4 Observed and calculated values for the general model (cross validation)

N	Observed Log1/*t*	Calculated Log1/*t*
1	3.523	5.855
2	4.523	6.146
3	5.398	6.195
4	6.523	6.078
5	6.398	6.078
6	8.301	5.806
7	6.699	5.948
8	7.000	6.126
9	6.398	6.087
10	5.955	5.856
11	6.046	6.618
12	7.155	8.011
13	7.222	8.46
14	8.398	8.258
15	9.000	8.139
16	8.699	7.684
17	6.921	7.516
18	6.699	6.24
19	8.398	7.803
20	7.398	7.863
21	9.000	8.688
22	9.921	8.466
23	9.155	8.172
24	7.699	7.977
25	7.097	6.703
26	9.097	8.754
27	7.699	9.773
28	10.097	9.432
29	8.699	9.138
30	7.222	9.629
31	6.398	6.994
32	8.155	8.867
33	8.000	9.827
34	9.921	9.653
35	10.301	9.585
36	10.699	9.482
37	10.155	9.445
38	10.585	9.288
39	10.222	9.172
40	7.398	8.911

Table 5 Regression equations

Series	Equations	R	Error	Cross validation "R"
Sample 1 (R_2=H)	$Log(1/t) = 3.0704 + 0.1488V(R_1) - 0.0013V(R_1)^2$	0.95	0.59	0.86
	$Log(1/t) = 0.8689 + 1.9734d_1 - 0.1373d_1^2$	**0.98**	0.37	0.83[a]
	$Log(1/t) = -0.3408 + 1.6723d_2 - 0.0858d_2^2$	0.97	0.50	0.86
Sample 2 (R_2=SC_2H_5)	$Log(1/t) = 5.8514 + 0.0971V(R_1) - 0.0008V(R_1)^2$	0.98	0.31	0.80
	$Log(1/t) = 4.7416 + 1.1159d_1 - 0.0739d_1^2$	**0.99**	0.13	0.94
	$Log(1/t) = 4.7078 + 0.8927d_2 - 0.0482d_2^2$	0.97	0.37	0.73
Sample 3 (R_2=SCH_3)	$Log(1/t) = 6.8246 + 0.1039V(R_1) - 0.0009V(R_1)^2$	**0.98**	0.28	0.73
	$Log(1/t) = 5.5745 + 1.1705d_1 - 0.0789_1^2$	0.98	0.33	0.95
	$Log(1/t) = 4.5654 + 1.1427d_2 - 0.0618d_2^2$	0.98	0.32	0.77
Sample 4 (R_2=OC_2H_5)	$Log(1/t) = 7.4024 + 0.107V(R_1) - 0.001V(R_1)^2$	**0.97**	0.43	0.62
	$Log(1/t) = 6.2307 + 1.1315d_1 - 0.0834d_1^2$	0.94	0.61	0.67
	$Log(1/t) = 5.1055 + 1.185d_2 - 0.0697d_2^2$	0.96	0.54	0.74
Sample 5 (R_2=OCH_3)	$Log(1/t) = 6.0973 + 0.1510V(R_1) - 0.0013V(R_1)^2$	0.96	0.51	0.86
	$Log(1/t) = 4.4841 + 1.6839d_1 - 0.1129d_1^2$	**0.98**	0.33	0.90
	$Log(1/t) = 2.7745 + 1.6678d_2 - 0.0886d_2^2$	0.97	0.46	0.96

[a]The last value is removed. When this value is taken into account, the R cross validation is close to 0.64

calculations. The medium value of the standard deviation associated with the calculated value is partly due to the deviation associated with the experimental determination of the olfactive threshold.

The efficiency of the model is illustrated by the good dispersion of the cloud of point in Fig. 2, which shows observed vs predicted values of Log(1/*t*). However, significant correlations (shown by Table 2 below) were detected between some variables. We then tried to eliminate the correlated variables and establish a new model. The final model, found to be in agreement with the constraints of the regression analysis, is the following:

$$Log(1/t) = -27.761 + 0.598DM - 48.013N_{1c} - 5.823N_{2c} - 0.002W \quad (2)$$

N=40, r=0.82, and s=0.801

The correlation matrix between variable descriptors for this model is presented in Table 3.

Table 6 Calculated Log(1/*t*) optima with the descriptors used

Series	[$V_{op}(R_1)$, $Log(1/t)_{op}$]		[d_{1op}, $Log(1/t)_{op}$]		[d_{2op}, $Log(1/t)_{op}$]	
Sample 1 (R_2=H)	57.23	**7.328**	7.186	**7.960**	9.745	**7.808**
Sample 2 (R_2=SC_2H_5)	62.00	**9.041**	7.550	**8.954**	9.260	**8.841**
Sample 3 (R_2=SCH_3)	57.72	**9.823**	7.417	**9.916**	9.245	**9.847**
Sample 4 (R_2=OC_2H_5)	53.50	**10.264**	6.783	**10.069**	8.501	**10.142**
Sample 5 (R_2=OCH_3)	58.07	**10.482**	7.457	**10.762**	9.411	**10.623**

To test the model obtained, we used the cross validation technique, which consists of removing one compound from the global sample and establishing a new model on the basis of the remaining 39 compounds. The new model obtained is used to test the molecule that was removed. The results obtained from this technique are presented in Table 4. The correlation coefficient between the observed and calculated values is near 0.76. This is a good value for cross-validation tests.

To enhance our understanding of the factors governing olfactive thresholds and to find optimum chemical structures, we tried to establish models between the olfactive threshold and variables on the basis of sets of molecules with the same molecular skeleton.

Analysis of subsets of closely similar molecules showed that there are some nonlinear relations between the olfactive threshold and each descriptor calculated. Based on the data analysis, we tested the correlation analysis for small subsets of molecules (Table 5). The models that we attempted are bilinear. These models are more efficient than the equation established for the global set of molecules, if one excludes the size of the sample.

For all samples, a cross-validation test led to good results, which allowed us to conclude that the established models are statistically significant, except for two models for sample 4. Interesting statistical models were obtained with different descriptors; however, the best models are those involving the $V(R_1)$, d_1, or d_2 descriptors. Such models can lead to values of $V(R_1)$, d_1, and d_2 corresponding to optimum chemical structures; the values of these descriptors corresponding to optimum Log(1/*t*) are given in Table 6.

Analysis of these optima shows that, following each descriptor [$V(R_1)$,d_1, and d_2], the olfactive threshold varies from one series to another and follows the following order:

$$R_2 = H < SC_2H_5 < SCH_3 < OC_2H_5 < OCH_3$$

Conclusions

The present study was conducted for a set of compounds that belong to pyrazine which have low olfactive thresholds. Statistical analysis of the global model obtained showed that it is possible to establish some correlations between the olfactive thresholds and physico-chemical parameters, mainly calculated by the molecular mechanics approach.

When the models were established for homogenous chemical subsets, we obtained more efficient equations and good statistical parameters. The equations obtained are bilinear, which allows us to calculate the optimum chemical structures with minimal olfactive thresholds. Thus, by using our equations, we will be able to design and test new molecules.

References

1. Herent MF, Collin S (1998) J Agric Food Chem 46:1975–1980
2. Zakarya D (1992) Nouv J Chim 16:1039–1042
3. Hideki M, Satoru M (1988) J Agric Food Chem 36:584–587
4. Satoru M, Hideki M (1988) J Agric Food Chem 36:1242–1247
5. Bondi A (1964) J Phys Chem 68:441–451
6. Wiener H (1947) J Phys Chem 69:17–20