JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Quantitative Structure−Activity Relationship Modeling of Alcohol, Ester, Aldehyde, and Ketone Flavor Thresholds in Beer from Molecular Features

YONGXI TAN AND KARL J. SIEBERT*

Department of Food Science and Technology, Cornell University, Geneva, New York 14456-0462

Three quantitative structure−activity relationship approaches—principal components regression, partial least-squares regression, and alternating conditional expectations—were used to investigate relationships between the flavor thresholds of 38 alcohols, 40 esters, 45 aldehydes, and 43 ketones in beer and their structures. Strong nonlinear relationships between the logarithm of the flavor threshold and four or five structure descriptors were found for each class of compounds ($R^2 = 0.920$, 0.937, 0.920, and 0.928 for alcohols, esters, aldehydes, and ketones, respectively). Simple nonlinear relationships between the alcohol, ester, and aldehyde thresholds and the numbers of hydrogen atoms in the molecules were also demonstrated.

KEYWORDS: Quantitative structure−activity relationships; partial least-squares regression; alternating conditional expectations; flavor threshold

INTRODUCTION

Many classes of compounds have been shown to play an important role in the flavor characteristics of foods and fragrances (*1*). The flavor impacts of many of the identified compounds have been assessed in various systems. Often those in food have been reported as flavor thresholds (*2*). It is of interest to investigate the relationships between the sensory perception of compounds and their molecular properties.

Many powerful computational methods, including artificial neural networks (ANN) (*3*), partial least-squares regression (PLSR) (*4*−*6*), and comparative molecular field analysis (CoMFA) (*7*), have been used in quantitative structure−activity relationship (QSAR) studies to investigate relationships between the properties or activities of compounds and their structures, especially in the fields of drug design and medicinal chemistry. QSAR techniques have also been used by flavor chemists to explore relationships between molecular structure and flavor properties and to give insight into the mechanism of interaction between flavor compounds and flavor receptors. Rossiter (*8*) investigated the correlation between chemical structure and the intensities of several fruity flavor characters of 27 aliphatic esters by three QSAR approaches: CoMFA, principal components analysis (PCA), and a Hansch approach. Multiple linear regression (MLR) and CoMFA were applied to model the aroma thresholds in water of 46 pyrazines with bell-pepper aroma (*9*). The flavor thresholds of organic acids in beer were modeled as a function of their molecular properties using principal components regression (PCR), MLR, and PLSR (*10*, *11*). More recently, ANN was used to classify the aroma types of 98 pyrazines and to predict their threshold values (*3*). ANN and self-organizing molecular field analysis were applied to analyze the aroma quality of pyrazine derivatives (*12*).

The objective of the work reported here was to investigate the relationships between the flavor thresholds of alcohols, esters, ketones, and aldehydes in beer and their structures using several QSAR approaches.

MATERIALS AND METHODS

Threshold Data and Structure Descriptors. Flavor thresholds in beer for four classes of flavor compounds commonly found in foods (38 alcohols, 40 esters, 45 aldehydes, and 43 ketones) were obtained from the literature (*13*, *14*) (see **Tables 1**−**4**). The data were expressed in milligrams per liter and were converted to millimoles per liter for modeling.

The variables initially chosen for modeling the flavor thresholds of the four classes of flavor compounds included physical and chemical property data (including boiling point, melting point, flash point, and density) and structure descriptors obtained from inspection of molecular structure. These included the number of carbon atoms in the molecule (C__NO), the number of hydrogen atoms (H__NO), the number of oxygen atoms (O__NO), the number of carbon−carbon double bonds (CC__NO), the number of conjugated double bonds (CONJ), the number of carbon atoms that connect to three or four non-hydrogen atoms (3C__NO), the number of carbon atoms on the alcohol side of an ester linkage (AC__NO), and the number of carbon atoms in the longer chain adjacent to a ketone function (LC__NO). Models were constructed with different combinations of these molecular and property descriptors. Leave-one-out cross-validation was performed. The models with the highest cross-validated multiple correlation coefficient squared (R^2cv) for each class were considered to be optimal combinations.

Computational Methods. Three QSAR approaches—PCR (*15*), PLSR (*4*−*6*), and alternating conditional expectations (ACE) (*16*−*19*)—were applied using the SCAN, Software for Chemometric Analysis, release 1.1 program (Minitab Inc., State College, PA). The relationships between molecular structure and log(flavor thresholds) for each of four classes of compounds were investigated. The modeling capability

* Author to whom correspondence should be addressed [telephone (315) 787-2299; fax (315) 787-2284; e-mail kjs3@cornell.edu].

10.1021/jf035149j CCC: $27.50
Published on Web 04/16/2004

Table 1. Flavor Thresholds Reported in Beer (*13*, *14*) and Structure Descriptors[a] Chosen for 38 Alcohols

	threshold								threshold						
alcohol	mg/L	mmol/L	C__NO	H__NO	CC__NO	CONJ	3C__NO	alcohol	mg/L	mmol/L	C__NO	H__NO	CC__NO	CONJ	3C__NO
methanol	10000	310	1	4	0	0	0	benzyl alcohol	900	8.3	7	8	3	3	1
ethanol	14000	300	2	6	0	0	0	1-hepten-3-ol	0.15	0.0013	7	14	1	0	0
1-propanol	800	13	3	8	0	0	0	1-heptanol	1	0.0086	7	16	0	0	0
2-propanol	1500	25	3	8	0	0	0	2-heptanol	0.25	0.0022	7	16	0	0	0
1-butanol	450	6.1	4	10	0	0	0	2-phenylethanol	125	1.02	8	10	3	3	1
2,3-butanediol	4500	50	4	10	0	0	0	tyrosol	200	1.4	8	10	3	3	1
2-butanol	16	0.22	4	10	0	0	0	1-octen-3-ol	0.2	0.0016	8	16	1	0	0
isobutanol	200	2.7	4	10	0	0	1	1-octanol	0.9	0.0069	8	18	0	0	0
tert-butanol	1600	22	4	10	0	0	1	2-octanol	0.04	0.00031	8	18	0	0	0
1-penten-3-ol	0.35	0.0041	5	10	1	0	0	1-nonanol	0.08	0.00055	9	20	0	0	0
1-pentanol	80	0.91	5	12	0	0	0	2-nonanol	0.075	0.00052	9	20	0	0	0
2-pentanol	45	0.51	5	12	0	0	0	linalool	0.08	0.00052	10	18	2	0	2
3-pentanol	50	0.57	5	12	0	0	0	nerol	0.5	0.0032	10	18	2	0	2
2-methyl-1-butanol	65	0.74	5	12	0	0	1	α-terpineol	2	0.013	10	18	1	0	3
isoamyl alcohol	70	0.79	5	12	0	0	1	1-decanol	0.18	0.0011	10	22	0	0	0
cis-3-hexen-1-ol	13	0.13	6	12	1	0	0	2-decanol	0.015	0.00009	10	22	0	0	0
trans-2-hexen-1-ol	15	0.15	6	12	1	0	0	1-undecanol	0.5	0.0029	11	24	0	0	0
1-hexanol	4	0.039	6	14	0	0	0	2-undecanol	0.07	0.00041	11	24	0	0	0
2-hexanol	4	0.039	6	14	0	0	0	1-dodecanol	0.4	0.0022	12	26	0	0	0

[a] C__NO, number of carbon atoms; H__NO, number of hydrogen atoms; CC__NO, number of carbon–carbon double bonds; CONJ, number of conjugated double bonds; 3C__NO, number of carbon atoms connecting to three or four non-hydrogen atoms.

Table 2. Flavor Thresholds Reported in Beer (*13*, *14*) and Structure Descriptors[a] Chosen for 40 Esters

	threshold								threshold						
ester	mg/L	mmol/L	C__NO	H__NO	O__NO	AC__NO	CONJ	ester	mg/L	mmol/L	C__NO	H__NO	O__NO	AC__NO	CONJ
methyl formate	5000	83	2	4	2	1	0	isoamyl propionate	0.7	0.0049	8	16	2	5	0
methyl acetate	550	7.4	3	6	2	1	0	*n*-hexyl acetate	3.5	0.024	8	16	2	6	0
ethyl formate	150	2.0	3	6	2	2	0	ethyl heptanoate	0.4	0.0025	9	18	2	2	0
ethyl acetate	30	0.34	4	8	2	2	0	*n*-amyl butyrate	0.6	0.0038	9	18	2	5	0
ethyl pyruvate	85	0.73	5	8	3	2	2	heptyl acetate	1.4	0.0088	9	18	2	7	0
ethyl lactate	250	2.1	5	10	3	2	0	2-phenylethyl acetate	3.8	0.023	10	12	2	8	3
n-propyl acetate	30	0.29	5	10	2	3	0	ethyl octanoate	0.9	0.0052	10	20	2	2	0
isobutyl formate	30	0.29	5	10	2	4	0	*n*-octyl acetate	0.5	0.0029	10	20	2	8	0
ethyl butyrate	0.4	0.0034	6	12	2	2	0	methyl caprate	1	0.0054	11	22	2	1	0
ethyl isobutyrate	5	0.043	6	12	2	2	0	ethyl nonanoate	1.2	0.0064	11	22	2	2	0
sec-butyl acetate	12	0.10	6	12	2	4	0	isoamyl hexanoate	0.9	0.0048	11	22	2	5	0
isobutyl acetate	1.6	0.014	6	12	2	4	0	ethyl decanoate	1.5	0.0075	12	24	2	2	0
n-butyl acetate	7.5	0.065	6	12	2	4	0	octyl butyrate	1.2	0.0060	12	24	2	8	0
tert-butyl acetate	24	0.21	6	12	2	4	0	ethyl undecanoate	1	0.0047	13	26	2	2	0
isoamyl formate	5	0.043	6	12	2	5	0	ethyl laurate	3.5	0.015	14	28	2	2	0
ethyl levulinate	300	2.1	7	12	3	2	0	isoamyl nonanoate	2	0.0088	14	28	2	5	0
ethyl isovalerate	1.3	0.010	7	14	2	2	0	octyl hexanoate	5	0.022	14	28	2	8	0
ethyl valerate	0.9	0.0069	7	14	2	2	0	ethyl palmitate	1.5	0.0053	18	36	2	2	0
isoamyl acetate	1.2	0.0092	7	14	2	5	0	ethyl linoleate	4	0.013	20	36	2	2	0
ethyl hexanoate	0.21	0.0015	8	16	2	2	0	ethyl oleate	3.5	0.011	20	38	2	2	0

[a] C__NO, number of carbon atoms; H__NO, number of hydrogen atoms; O__NO, number of oxygen atoms; AC__NO, number of carbon atoms at the alcohol side of ester linkage; CONJ, number of conjugated double bonds.

(goodness of fit) was judged by the residual sum of squares (residual SS) and by the multiple correlation coefficient squared, R^2. The prediction capability (goodness of prediction) was indicated by the residual predictive error sum of squares (residual PRESS) and by the cross-validated R^2 (R^2cv).

ACE graphs of the transformations of the predictor variables selected as optimal were examined.

RESULTS AND DISCUSSION

The sensory data used (*13*, *14*) were obtained with the ascending method of limits (*20*), and compounds were purified to constant threshold. Panelists used smell and/or taste as they saw fit to distinguish the samples with added substances from the control samples. Individual flavor thresholds were determined for each compound using a moderate-size panel (typically $\geq$12 individuals). The panel threshold was computed as the geometric average of the individual thresholds.

Principal components regression attempts to find orthogonal components from the predictors (*X* data matrix) such that the first principal component shows the largest variation in *X* and the second principal component orthogonal to the first represents the second largest variation, etc. (*15*). Then the response is related to the principal components by ordinary least squares regression (OLS). PCR is advantageous when a data set is overdetermined (many variables and few observations) or highly collinear, which is often the case in QSAR and spectral analysis. One drawback of PCR is inefficiency, because the components are selected on the basis of their ability to explain variance in the *X* matrix rather than for their ability to predict the response.

Partial least-squares regression is frequently used in QSAR to model the relation between two data matrices, **X** and **Y** (*6*). To overcome the shortcoming of PCR, PLSR attempts to extract a set of orthogonal components from the **X** matrix that are of importance in predicting the response. The main attraction of this biased regression method is its capability to handle multicollinear data sets, where the unbiased least-squares solution is not reliable due to large variance.

Table 3. Flavor Thresholds Reported in Beer (*13*, *14*) and Structure Descriptors[a] for 45 Aldehydes

	threshold							threshold					
aldehyde	mg/L	mmol/L	H__NO	O__NO	CC__NO	3C__NO	aldehyde	mg/L	mmol/L	H__NO	O__NO	CC__NO	3C__NO
formaldehyde	400	13	2	1	0	0	hydrocinnamaldehyde	1	0.007	10	1	3	1
glyoxal	7000	120	2	2	0	0	hexanal	0.35	0.003	12	1	0	0
glyoxylic acid	2000	27	2	3	0	1	2-ethylbutanal	6	0.06	12	1	0	1
acetaldehyde	25	0.57	4	1	0	0	*cis*-4-heptenal	0.0004	3.6E−06	12	1	1	0
acrolein	15	0.27	4	1	1	0	cuminaldehyde	0.4	0.0027	12	1	3	3
furfural	150	1.6	4	2	2	1	heptanal	0.08	0.00070	14	1	0	0
propionaldehde	30	0.52	6	1	0	0	2-ethyl-2-hexenal	0.2	0.0016	14	1	1	1
D-(+)-glyceraldehyde	125	1.4	6	3	0	1	*trans*-2,*cis*-6-nonadienal	0.00005	3.6E−07	14	1	2	0
crotonal	8	0.11	6	1	1	0	*trans*-2,*trans*-4-nonadienal	0.0003	2.2E−06	14	1	2	0
5-(hydroxymethyl)furfural	1000	7.9	6	3	2	2	octanal	0.04	0.00031	16	1	0	0
5-methylfurfural	20	0.18	6	2	2	2	2-ethylhexanal	1	0.0078	16	1	0	1
benzaldehyde	2	0.019	6	1	3	1	*trans*-2-nonenal	0.00011	7.8E−07	16	1	1	0
butyraldehyde	1	0.014	8	1	0	0	*trans*-2,*trans*-4-decadienal	0.0003	2.0E−06	16	1	2	0
methional	0.25	0.002	8	1	0	0	citral	0.15	0.00099	16	1	2	2
aldol	8	0.091	8	2	0	1	nonanal	0.02	0.00014	18	1	0	0
isobutanal	1	0.014	8	1	0	1	*trans*-2-decenal	0.001	6.5E−06	18	1	1	0
trans-2,*trans*-4-hexadienal	0.8	0.008	8	1	2	0	citronellal	4	0.026	18	1	1	2
phenylacetaldehyde	1.6	0.013	8	1	3	1	decanal	0.006	3.8E−05	20	1	0	0
cinnamaldehyde	6	0.045	8	1	4	1	hydroxycitronellal	1.5	0.0087	20	2	0	2
pentanal	0.5	0.006	10	1	0	0	10-undecenal	0.0035	2.1E−05	20	1	1	0
2-methylbutanal	1.25	0.015	10	1	0	1	undecanal	0.0035	2.1E−05	22	1	0	0
isopentanal	0.6	0.007	10	1	0	1	dodecanal	0.004	2.2E−05	24	1	0	0
trans-2-hexenal	0.6	0.006	10	1	1	0							

[a] H__NO, number of hydrogen atoms; O__NO, number of oxygen atoms; CC__NO, number of carbon–carbon double bonds; 3C__NO, number of carbon atoms connecting to three or four non-hydrogen atoms.

Alternating conditional expectations is a nonlinear, nonparametric regression method that estimates optimal nonlinear transformations of both response and predictor variables (*16*, *19*). In regression problems, the use of transformations of independent or dependent variables is a common practice to aid in understanding nonlinear relationships between predictors and response. Such transformations are often arbitrary or intuition-inspired and generally are not the optimal ones. ACE provides an approach to find transformations that produce the best-fitting additive model through an iterative algorithm using conditional expectations (*16*). Because knowledge about the shape of the transformations obtained in ACE is helpful in understanding the relationships between structure and activity, ACE has recently been applied in QSAR studies (*18*, *21*, *22*).

In the SCAN implementation of ACE, this nonlinear regression model has the form

$$y = \Sigma_i f_i(x_i) + e \quad (1)$$

where the transform functions f_i are smooth but otherwise unrestricted functions of the predictor variables, which can be obtained by smoothers estimated using an iterative least-squares algorithm. The response is then modeled as the sum of the transform functions. One of the advantages of ACE over PCR and PLSR is its capability to model nonlinear relationships between predictors and response. In general, this model has no simple analytical form and the ACE predictions in SCAN include two steps: First, the transform function values corresponding to each predictor are calculated by linear interpolation between two values according to the estimated transform function [x_i, f_i (x_i)]. Then, the predicted response values are obtained as the sum of transform function values.

A method to represent the degree of branching of a molecule was devised; the number of carbon atoms with three or four of its bonds connected to atoms other than hydrogen was defined as the 3C__NO.

Models were constructed with various combinations of the structure descriptors and the chemical properties. The strongest models (highest R^2cv) for each of the four classes of compounds studied were obtained with combinations of four or five structure descriptors (shown in **Tables 1**−**4**) and without any of the property data.

The relationships between the logarithms of thresholds (expressed as millimoles per liter) of 38 alcohols and their structure descriptors (**Table 1**) were investigated using PCR, PLSR, and ACE. The results are summarized in **Table 5**. The optimal number of PLSR components derived from the five predictors was four (corresponding to the lowest residual PRESS and the largest R^2cv), which was the same as the optimal number of PCR components. This result suggests that the relationship can be modeled with a modest number of fundamental predictors. It was obvious that ACE gave better results than the other two approaches, with $R^2 = 0.920$ and $R^2\text{cv} = 0.826$. The R^2cv is more conservative than R^2 and generally considered to provide a more realistic estimate of the ability of a model to make predictions for other conditions (*23*). The notably stronger model produced with ACE indicated that one or more of the relationships between the structure descriptors (C__NO, H__NO, CC__NO, CONJ, and 3C__NO) and the logarithms of alcohol thresholds were nonlinear. The ACE transformations are shown in **Figure 1**; this revealed the underlying relationships between the predictors and response. **Figure 2** shows the relationship between the logarithms of observed alcohol thresholds and the predictions obtained using ACE. Three compounds deviated somewhat from the model; these were 1-penten-3-ol and 2-butanol (above the line) and 2,3-butanediol (below). Examining the forms of the transforms in **Figure 1** permits interpretation of the effects of the predictors on the flavor thresholds of alcohols. Higher flavor thresholds (weaker impacts) were observed with smaller numbers of carbon atoms, smaller numbers of carbon−carbon double bonds, larger numbers of conjugated double bonds, and a greater extent of branching. Examination of **Table 1** shows that only three compounds had the higher level of conjugated double bonds, and all three were aromatic, so the effect here may be mainly differentiation between aliphatic and aromatic alcohols. Obviously aromatic compounds differ to a much greater extent than is represented by three conjugated double bonds; however,

Table 4. Flavor Thresholds Reported in Beer (*13*, *14*) and Structure Descriptors[a] for 43 Ketones

ketone	threshold		C_NO	H_NO	O_NO	CONJ	3C_NO
	mg/L	mmol/L					
acetone	200	3.4	3	6	1	0	1
pyruvic acid	300	3.4	3	4	3	2	2
2-butanone	80	1.1	4	8	1	0	1
acetoin	50	0.57	4	8	2	0	2
diacetyl	0.15	0.0017	4	6	2	2	2
oxalacetic acid	500	3.8	4	4	5	2	3
2-pentanone	30	0.35	5	10	1	0	1
3-hydroxy-3-methyl-2-butanone	400	3.9	5	10	2	0	2
3-methyl-2-butanone	60	0.70	5	10	1	0	2
3-pentanone	30	0.35	5	10	1	0	1
1-penten-3-one	0.03	0.00036	5	8	1	2	1
2,3-pentanedione	0.9	0.0090	5	8	2	2	2
cyclopentanone	200	2.4	5	8	1	0	1
2-hexanone	4	0.040	6	12	1	0	1
3,3-dimethyl-2-butanone	25	0.25	6	12	1	0	2
4-methyl-2-pentanone	5	0.050	6	12	1	0	2
2,3-hexanedione	15	0.13	6	10	2	2	2
cyclohexanone	40	0.41	6	10	1	0	1
mesityl oxide	4	0.041	6	10	1	2	2
2-acetylfuran	80	0.73	6	6	2	3	2
2,4-dimethyl-3-pentanone	8	0.070	7	14	1	0	3
2-heptanone	2	0.018	7	14	1	0	1
3-heptanone	3	0.026	7	14	1	0	1
4-heptanone	4	0.035	7	14	1	0	1
5-methyl-2-hexanone	7	0.061	7	14	1	0	2
4-methylcyclohexanone	25	0.22	7	12	1	0	2
2-octanone	0.25	0.0019	8	16	1	0	1
3-octanone	0.5	0.0039	8	16	1	0	1
6-methyl-3-heptanone	1.2	0.0094	8	16	1	0	2
1-octen-3-one	0.000025	2.0E–07	8	14	1	2	1
o-aminoacetophenone	0.005	3.7E–05	8	9	1	4	3
acetophenone	3	0.025	8	8	1	4	2
2,6-dimethyl-4-heptanone	8	0.056	9	18	1	0	3
2-nonanone	0.2	0.0014	9	18	1	0	1
2-decanone	0.25	0.0016	10	20	1	0	1
3-decanone	0.3	0.0019	10	20	1	0	1
d-(+)-carvone	0.4	0.0027	10	14	1	2	4
benzylacetone	2.5	0.017	10	12	1	3	2
trans-4-phenyl-3-buten-2-one	1	0.0068	10	10	1	5	2
2-undecanone	0.4	0.0023	11	22	1	0	1
2-dodecanone	0.25	0.0014	12	24	1	0	1
α-ionone	0.0026	1.4E–05	13	20	1	2	4
β-ionone	0.0013	6.8E–06	13	20	1	3	4

[a] C_NO, number of carbon atoms; H_NO, number of hydrogen atoms; O_NO, number of oxygen atoms; CONJ, number of conjugated double bonds; 3C_NO, number of carbon atoms connecting to three or four non-hydrogen atoms.

Table 5. Results of QSAR Modeling of Alcohol Threshold Data in **Table 1** Using PCR, PLSR, and ACE

method	components	residual SS[a]	R^2	residual PRESS[b]	R^2cv[c]
PCR	4	20.62	0.835	26.49	0.788
PLSR	4	20.62	0.835	26.49	0.788
ACE	5	10.02	0.920	21.76	0.826

[a] Residual sum of squares. [b] Predicted residual error sum of squares. [c] R^2cv = cross-validated R^2.

this appears to suffice in this modeling approach because the aromatic compounds do not appear as outliers, except in the ketone models. The effect of increasing numbers of hydrogen atoms (see **Figure 1**) was an overall trend to increase the threshold, particularly with the very highest numbers of hydrogen atoms. There was a distinct peak in threshold for alcohols with 12 hydrogen atoms. Because the number of hydrogen atoms in a molecule tends to increase with the number of carbon atoms, the fact that smaller carbon numbers and larger hydrogen numbers were both associated with higher thresholds appears to be contradictory. Meilgaard (*13*) observed that the thresholds of several homologous series of compounds (expressed as milligrams per liter) versus the number of carbons in a molecule in nearly every case declined to a minimum (often near eight or nine carbons) and then rose again to form a "v" or "u" shape. In the case of the primary, straight-chain, saturated alcohols, this pattern can be seen with a minimum threshold (here expressed in millimoles per liter) at nine carbons (see the data points in **Figure 3**). In this series of compounds the contributions of the predictors other than C_NO and H_NO are all constant. It can be seen that the thresholds of this compound subset are predicted reasonably well (the line in **Figure 3**) because of the increasingly steep contribution of H_NO. Implicit in either C_NO or H_NO should be information about molecular size and polarity. Stronger sensory impacts (lower thresholds) are associated with larger numbers of carbon atoms (larger molecules) and carbon−carbon double bonds and with smaller numbers of hydrogen atoms (smaller molecules or less saturation) and conjugated double bonds (lower saturation or in this case aliphatic rather than aromatic nature) and less branching.

Table 6 shows the results of QSAR modeling of the 40 ester thresholds in **Table 2** using PCR, PLSR, and ACE. The results

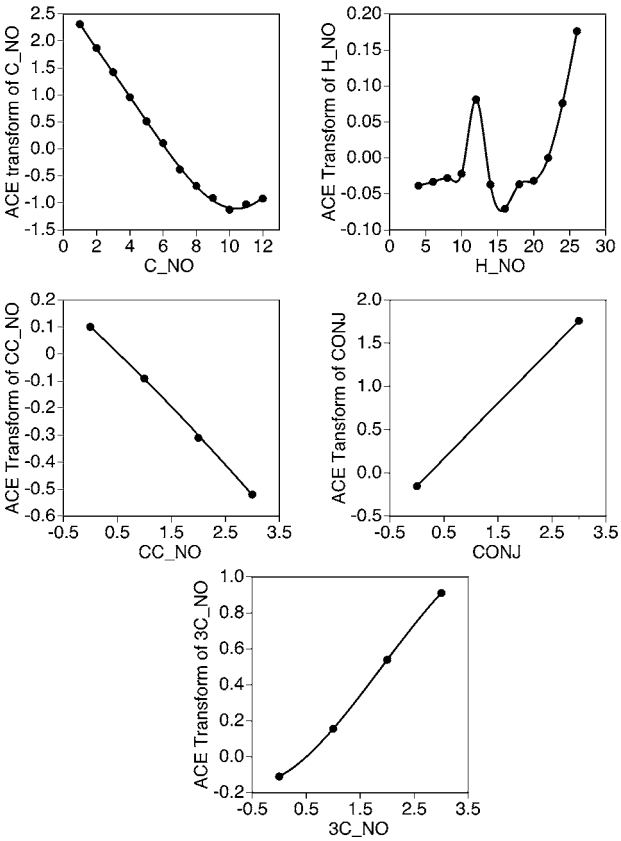

Figure 1. Transforms applied to C_NO, H_NO, CC_NO, CONJ, and 3C_NO in the ACE model of alcohol flavor thresholds.

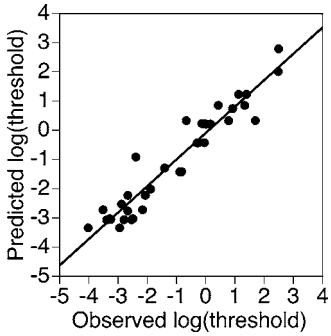

Figure 2. Relationship ($R^2 = 0.920$) between the logarithms of observed alcohol thresholds (mmol/L) and predictions obtained using ACE fitting.

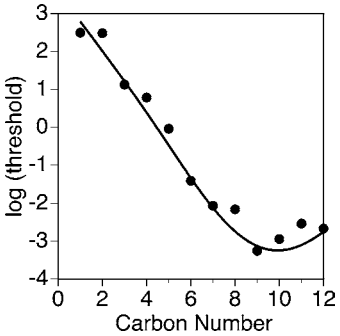

Figure 3. Flavor thresholds of saturated, unbranched, primary alcohols as a function of carbon numbers (data points) and thresholds predicted from the ACE model (line).

from ACE ($R^2 = 0.937$ and R^2cv = 0.873) were much better than those from PCR and PLSR (which had best prediction models with $R^2 = 0.491$ and 0.484, respectively; both had R^2-cv = 0.379). The low R^2 and R^2cv obtained with PCR and PLSR indicated very poor performance in modeling and prediction of ester thresholds. Again, this occurred because the actual underlying relationships between the five structure descriptors used and the logarithms of ester thresholds were nonlinear (see **Figure 4**), and both PCR and PLSR are based on fitting linear relationships between response and predictors. The relationship between the logarithms of thresholds and the ACE model is displayed in **Figure 5**. The compounds that deviated most from the model here are ethyl butyrate and isobutyl butyrate (both above) and *n*-hexyl acetate and *tert*-butyl acetate (both below). For esters, higher flavor thresholds (weaker sensory impacts) were generally associated with larger numbers of carbon and oxygen atoms (**Figure 4)** and with smaller numbers of hydrogen atoms and conjugated double bonds. In this case the downward impact on thresholds of increasing H_NO was greater than the upward effect of increasing C_NO. Increasing the carbon number of the ester alcohol chain (AC_NO) tended to moderately decrease the threshold; there were, however, two large exceptions to this trend. The first was an extremely low point in thresholds at AC_NO = 2 (corresponding to ethyl esters). The second was a high threshold value at AC_NO = 4 (corresponding to butyl esters). Stronger ester sensory impacts

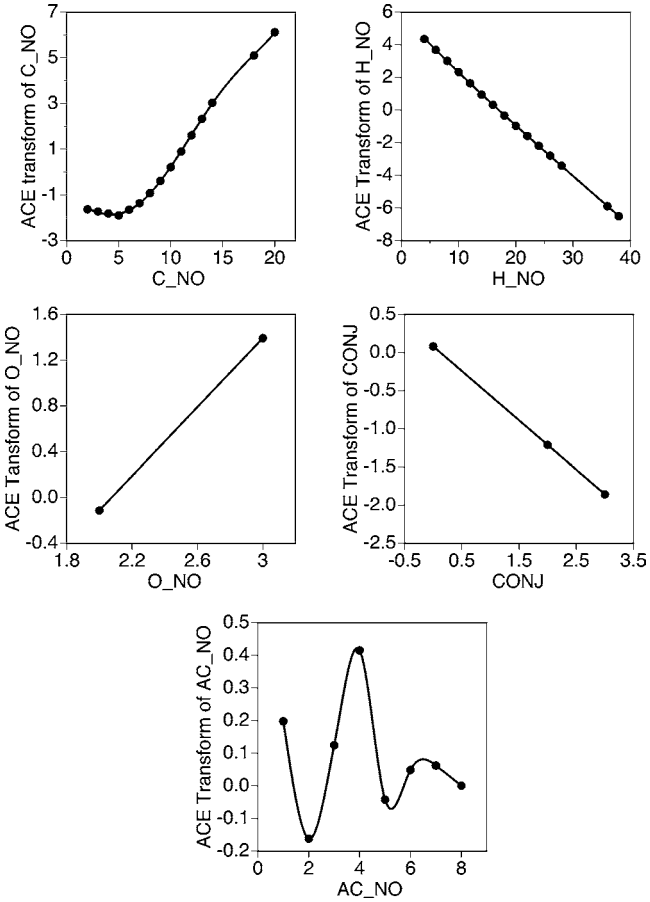

Figure 4. Transforms applied to C_NO, H_NO, O_NO, CONJ, and AC_NO in the ACE model of ester flavor thresholds.

Table 6. Results of QSAR Modeling of Ester Threshold Data in **Table 2** Using PCR, PLSR, and ACE

method	components	residual SS[a]	R^2	residual PRESS[b]	R^2cv[c]
PCR	3	23.59	0.491	28.81	0.379
PLSR	1	23.95	0.484	28.79	0.379
ACE	5	2.90	0.937	5.89	0.873

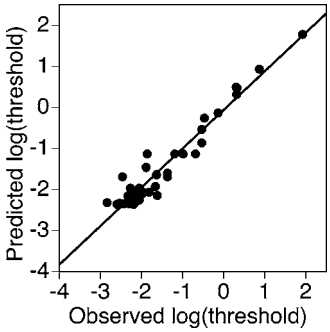

Figure 5. Relationship ($R^2 = 0.937$) between the logarithms of observed ester thresholds (mmol/L) and those predicted using ACE fitting.

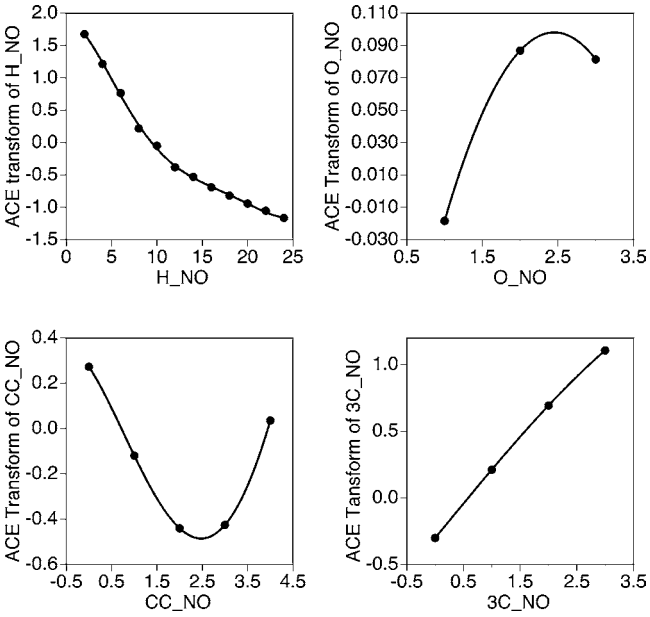

Figure 6. Transforms applied to H_NO, O_NO, CC_NO, and 3C_NO in the ACE model of aldehyde flavor thresholds.

Table 7. Results of QSAR Modeling of Aldehyde Threshold Data in **Table 3** Using PCR, PLSR. and ACE

method	components	residual SS[a]	R^2	residual PRESS[b]	R^2cv[c]
PCR	4	36.70	0.804	46.97	0.749
PLSR	4	36.70	0.804	46.97	0.749
ACE	4	15.02	0.920	34.09	0.818

(lower thresholds) then were associated with smaller numbers of carbon atoms (smaller molecules) and oxygen atoms and with larger numbers of hydrogen atoms (larger molecules or greater saturation) and conjugated double bonds (lower saturation). The effect of AC_NO varied as described above.

QSAR models of the 45 aldehyde thresholds in **Table 3** were obtained using PCR, PLSR, and ACE, as shown in **Table 7**. ACE again performed better than PCR or PLSR. The ACE transformation plots for each structure descriptor are shown in **Figure 6**. **Figure 7** shows the overall ACE model. The compounds with the largest residuals were *cis*-4-heptenal and glyoxylic acid (above) and 10-undecanal (below). The effects of the predictors can be seen from the transform plots in **Figure 6**. For aldehydes, higher flavor thresholds (weaker sensory impacts) were generally associated with higher branching (3C_NO), smaller numbers of hydrogen atoms, and larger numbers of oxygen atoms. Higher thresholds were also associated with the lowest (none) and highest (four) numbers of carbon−carbon double bonds. Examination of **Table 3** shows

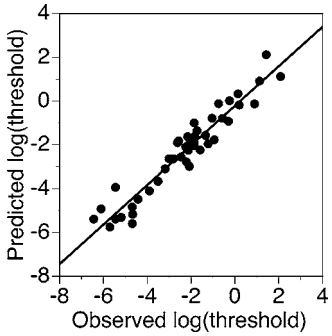

Figure 7. Relationship ($R^2 = 0.920$) between the logarithms of observed aldehyde thresholds (mmol/L) and those predicted using ACE fitting.

Table 8. Results of QSAR Modeling of Ketone Threshold Data in **Table 4** Using PCR, PLSR, and ACE

method	components	residual SS[a]	R^2	residual PRESS[b]	R^2cv[c]
PCR	5	33.27	0.700	52.90	0.523
PLSR	5	33.27	0.700	52.90	0.523
ACE	5	8.00	0.928	42.28	0.619

that only five compounds had either of the two higher levels of carbon−carbon double bonds, and all five of these were aromatic, so the effect here may be mainly a differentiation between aliphatic and aromatic aldehydes. Only one compound, cuminaldehyde, had 3C_NO = 3. Stronger sensory impacts (lower thresholds) would conversely be associated with small 3C_NO values and numbers of oxygen atoms and with larger numbers of hydrogen atoms (larger molecules or greater saturation). For the aliphatic aldehydes a stronger flavor impact was associated with larger numbers of carbon−carbon double bonds, whereas the direction was opposite for aromatic aldehydes.

QSAR models of the 43 ketone thresholds in **Table 4** were constructed using PCR, PLSR, and ACE, and the results are shown in **Table 8**. As before, ACE provided better performance than PCR or PLSR. The R^2cv for the ACE ketone model was lower than with ACE models of the other compound classes. The ACE transformation plots for each structure descriptor are shown in **Figure 8**, whereas the relationship between the logarithms of observed thresholds and the model obtained using ACE is shown in **Figure 9**. The compounds with relatively large residuals were *o*-aminoacetophenone, diacetyl, and cyclohexanone (all above) and acetophenone, mesityl oxide, 2,3-hexanedione, and 2-butanone (below). For ketones higher flavor thresholds (weaker sensory impacts) were generally associated with larger numbers of carbon and oxygen atoms and with smaller numbers of hydrogen atoms and conjugated double bonds. It is interesting that the number of oxygen atoms fits a linear function even though the compounds with different numbers of oxygen atoms varied considerably in their functional groups (including diketones and carboxy and hydroxy groups). The effect of branching was that molecules with modest amounts of branching tended to have higher thresholds. Stronger sensory impacts (lower thresholds) were conversely associated with smaller numbers of carbon and oxygen atoms, with higher numbers of hydrogen atoms (larger molecules or greater saturation) and conjugated double bonds, and with either no or much branching.

The predictor variable importance values given by direct ACE modeling are shown in **Table 9**. These magnitudes show the relative importance of the different predictors. The only term common to all four models was H_NO, and this was the most

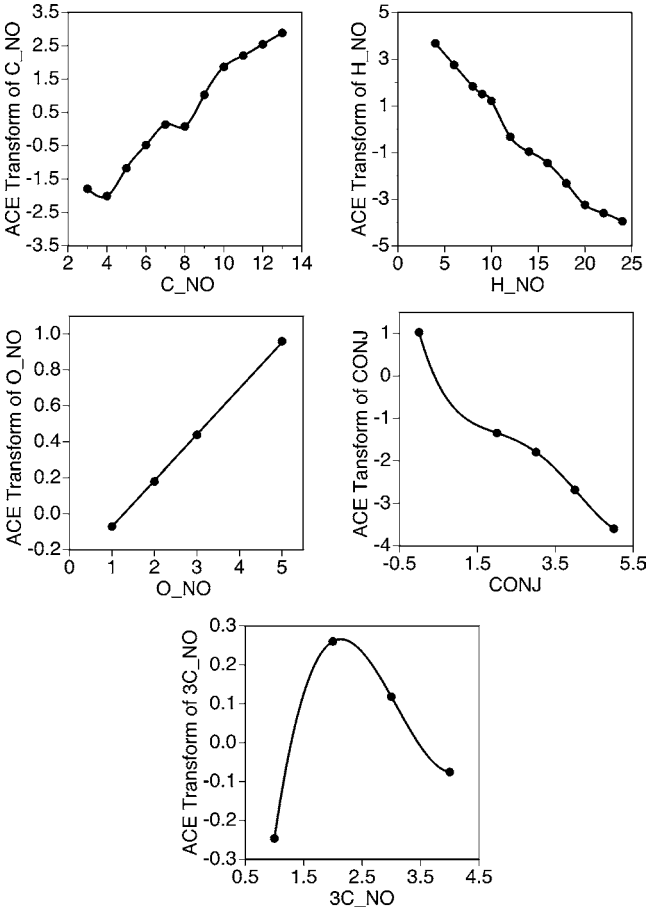

Figure 8. Transforms applied to C__NO, H__NO, O__NO, CONJ, and 3C__NO in the ACE model of ketone flavor thresholds.

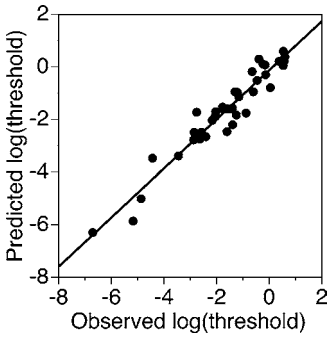

Figure 9. Relationship ($R^2 = 0.928$) between the logarithms of observed ketone thresholds (mmol/L) and the predictions from ACE fitting.

Table 9. Predictor Importances for ACE Models of Alcohol, Ester, Aldehyde, and Ketone Flavor Thresholds

class	C__NO	H__NO	O__NO	CC__NO	CONJ	3C__NO	AC__NO
alcohols	0.944	0.058		0.189	0.522	0.228	
esters	2.25	2.69	0.185		0.402		0.364
aldehydes		0.796	0.040	0.306		0.385	
ketones	1.39	2.06	0.188		1.41	0.238	

influential term for all but alcohols. It was of interest to evaluate the possibility of predicting the threshold of each class from this single term using ACE. This was attempted, and the results are shown in **Table 10**. Results with alcohols, esters, and aldehydes were quite respectable, although weaker in each case than the models with more terms already described. The ketone model was quite weak and unsatisfactory.

Table 10. Results of QSAR Modeling of the log(Flavor Threshold) as a Function of the Number of Hydrogen Atoms in the Molecules Using ACE

compound	residual SS	R^2	residual PRESS	R^2cv
alcohols	18.20	0.854	27.41	0.781
esters	7.650	0.835	9.593	0.793
aldehydes	56.28	0.699	69.44	0.628
ketones	65.49	0.410	91.46	0.176

Clearly it is possible to construct nonlinear QSAR models that relate the flavor threshold of a compound to its molecular features. This was successful for each of the four classes of compounds for which this was attempted. It was also possible to make somewhat weaker models that relate the flavor threshold to only the number of hydrogen atoms in the compounds for alcohols, esters, and aldehydes, but not ketones.

LITERATURE CITED

(1) *Allured's Flavor and Fragrance Materials—1996*; Allured Publishing: Carol Stream, IL, 1996.

(2) Lawless, H. T.; Heymann, H. *Sensory Evaluation of Food: Principles and Practices*; Kluwer Academic/Plenum Publishers: New York, 1998.

(3) Wailzer, B.; Klocker, J.; Buchbauer, G.; Ecker, G.; Wolschann, P. Prediction of the aroma quality and the threshold values of some pyrazines using artificial neural networks. *J. Med. Chem.* **2001**, *44*, 2805−2813.

(4) Wold, S.; Ruhe, A.; Wold, H.; Dunn, I. W. J. The collinearity problem in linear regression: the partial least squares approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735−743.

(5) Hoskulddsson, A. PLS regression methods. *J. Chemom.* **1988**, *2*, 211−228.

(6) Geladi, P.; Kowalski, B. R. Partial least squares regression (PLS): a tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.

(7) Cramer, R. D.; 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*, 5959−5967.

(8) Rossiter, K. J. Quantitative relationships between structure and the fruity odor of esters. *Perfum. Flavor.* **1996**, *21*, 33−46.

(9) Buchbauer, G.; Klein, C. T.; Wailzer, B.; Wolschann, P. Threshold-based structure−activity relationships of pyrazines with bell-pepper flavor. *J. Agric. Food Chem.* **2000**, *48*, 4273−4278.

(10) Siebert, K. J. Modeling the flavor thresholds of organic acids in beer as a function of their molecular properties. *Food Qual. Pref.* **1999**, *10*, 129−137.

(11) Siebert, K. J. Evaluation of a model of organic acid flavor thresholds in beer. *J. Am. Soc. Brew. Chem.* **2000**, *58*, 94−96.

(12) Klocker, J.; Wailzer, B.; Buchbauer, G.; Wolschann, P. Aroma quality differentiation of pyrazine derivatives using self-organizing molecular field analysis and artificial neural network. *J. Agric. Food Chem.* **2002**, *50*, 4069−4075.

(13) Meilgaard, M. C. Flavor chemistry of beer. I. Flavor interaction between principal volatiles. *Tech. Q. Master Brew. Assoc. Am.* **1975**, *12*, 107−117.

(14) Meilgaard, M. C. Flavor chemistry of beer. II. Flavor and threshold of 239 aroma volatiles. *Tech. Q. Master Brew. Assoc. Am.* **1975**, *12*, 151−168.

(15) Jackson, J. E. *User's Guide to Principal Components*; Wiley: New York, 1991.

(16) Breiman, L.; Friedman, J. H. Estimating optimal transformations for multiple regression and correlation. *J. Am. Stat. Assoc.* **1985**, *80*, 580−589.

(17) Blanco, M.; Coello, J.; Masdoch, S.; Puigdomènech, A. R. Modeling of an environmental parameter by use of the alternating conditional expectation method. *Chemom. Intell. Lab. Syst.* **1999**, *46*, 31−39.

(18) Clare, B. W. Structure−activity correlations for psychotomimetrics: 2. Phenylalkylamines: a treatment of nonlinearity using the alternating conditional expectations technique. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 71−92.

(19) Frank, I. E.; Lanteri, S. ACE: A non-linear regression model. *Chemom. Intell. Lab. Syst.* **1988**, *3*, 301−313.

(20) ASTM. *Standard Practice E679, Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits*; American Society for Testing and Materials: Philadelphia, PA, 1979.

(21) Nguyen-Cong, V.; Rode, B. M. Quantitative electronic structure−activity relationships of pyridinium cephalosphorins using nonparametric regression methods. *Eur. J. Med. Chem.* **1996**, *31*, 479−484.

(22) Nguyen-Cong, V.; Van Dang, G.; Rode, B. M. Using multivariate adaptive regression splines to QSAR studies of dihydroartemisinin derivatives. *Eur. J. Med. Chem.* **1996**, *31*, 797−803.

(23) Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part A*; Elsevier Science: Amsterdam, The Netherlands, 1997; Vol. 20A.

Received for review October 7, 2003. Revised manuscript received February 27, 2004. Accepted March 6, 2004. This study was supported by Robertet Flavors, Inc., Piscataway, NJ.

JF035149J