Quantitative Structure−Activity Relationship Studies of Sulfamates $RNHSO_3Na$: Distinction between Sweet, Sweet-Bitter, and Bitter Molecules

Michael G. B. Drew* and Gareth R. H. Wilden

Department of Chemistry, The University, Whiteknights, Reading, United Kingdom

William J. Spillane,* Rachel M. Walsh, Caroline A. Ryder, and John M. Simmie

Department of Chemistry, National University of Ireland, Galway, Ireland

Several different QSAR techniques have been applied to sweetness data for 50 sulfamates, $RNHSO_3Na$ (21 sweet, 20 sweet-bitter, and 9 bitter). Stepwise discriminant analysis has been used to separate the 50 molecules into 3 classes, sweet, sweet-bitter, and bitter. Cluster analysis using two principal components can clearly distinguish between the sweet and sweet-bitter molecules but not between all three classes. Regression analysis has been used to develop equations for parameters fitting to log(RS) (RS, relative sweetness). The genetic algorithm method has been used to select parameters, and high correlations between log(RS) and a range of parameters have been achieved. Molecular field analysis followed by selection of relevant grid points by genetic algorithm yielded a result in which six grid points gave a high correlation coefficient ($r^2 = 0.958$, $XVr^2 = 0.902$).

Keywords: *Sweetness; QSAR; sulfamates; bitterness*

INTRODUCTION

Quantitative structure−activity relationships (QSARs) have been used for many years to establish which structural and physical properties of a set of molecules are most responsible for their biological activity. The main thrust of QSAR studies has been in the field of drug design, but there have been several applications to the taste properties of molecules, particularly involving sweetness. The main groups of sweeteners that have been subjected to various types of QSAR studies are the nitro- and cyanoanilines (Deutsch et al., 1966; McFarland, 1971; Holtje and Kier, 1974; Iwamura, 1980; Kier, 1985; van der Heijden et al., 1985a), various aspartyl dipeptide derivatives (van der Heijden et al., 1979, 1985a; Iwamura, 1981; Miyashita et al., 1986a), oximes (perillartines) (Iwamura, 1980; van der Heijden et al., 1985b; Acton et al., 1976; Takahashi et al., 1982, 1984; Zalewski, 1992), β-(3-hydroxy and 4-methoxyphenylethylbenzenes (Miyashita et al., 1989), acesulfames, tryptophans, saccharins, chlorocarbohydrates, ureas, isocoumarins (van der Heijden, 1985a,b), and sulfamates (Spillane and McGlinchey, 1981; Spillane and Sheahan, 1989; Spillane et al., 1983, 1989, 1994; Miyashita et al., 1986b; Okuyama et al., 1988). For the latter group, which includes cyclamate (*N*-cyclohexylsulfamate), we have previously developed semi-QSARs using parameters measured with Corey−Pauling−Koltun (CPK) space-filling models of the R section of the sulfamates $RNHSO_3Na$. This approach has been successful for carbosulfamates (Spillane and McGlinchey, 1981; Spillane and Sheahan, 1989) and monosubstituted aromatic sulfamates (Spillane et al., 1989, 1994) and has shown good predictive ability. A pattern recognition method (Miyashita et al., 1986a,b; Ikuyama et al., 1988), discriminant analysis and principal component analysis (PCA) (Zalewski, 1992), and use of STERIMOL parameters (van der Heijden et al., 1985a,b) have been employed to develop structure−taste relationships for the carbosulfamates. Linear discriminant analysis successfully classified ~50 heterosulfamates into sweet and nonsweet categories with an overall classification of 86% (Spillane et al., 1983, 1989).

In the present work several different QSAR techniques have been applied to a database of relative sweetness (RS) for a series of 21 sodium sulfamates, $RNHSO_3Na$, of widely differing structural types, where R is straight and branched aliphatic, alicyclic, aromatic, and heterocyclic. In addition, another 20 sodium sulfamates (9 disubstituted aromatic and 11 monosubstituted aromatic), which displayed predominantly bitter and sweet components in their taste profiles, have been synthesized and are used in the present QSAR studies. A further 9 mono- and disubstituted benzosulfamates with exclusively bitter taste components are also included in this study. All molecules are shown in Figure 1; the sweet-tasting molecules are numbered **1**−**21**, the sweet-bitter-tasting molecules **22**−**41**, and the bitter molecules **42**−**50**.

Chemistry. All of the sulfamates were synthesized either by reaction of the appropriate amine with chlorosulfonic acid in chloroform (Spillane et al., 1993) or by reaction of amine with pyridine− or α-picoline−sulfur trioxide in excess base (Spillane et al., 1993) as solvent. Sodium 2-thiazolylsulfamate was prepared according to the procedure of Hurd and Kharasch (1946). The

* Author to whom correspondence should be addressed (e-mail m.g.b.drew@reading.ac.uk).

S0021-8561(98)00095-8 CCC: $15.00
Published on Web 07/11/1998

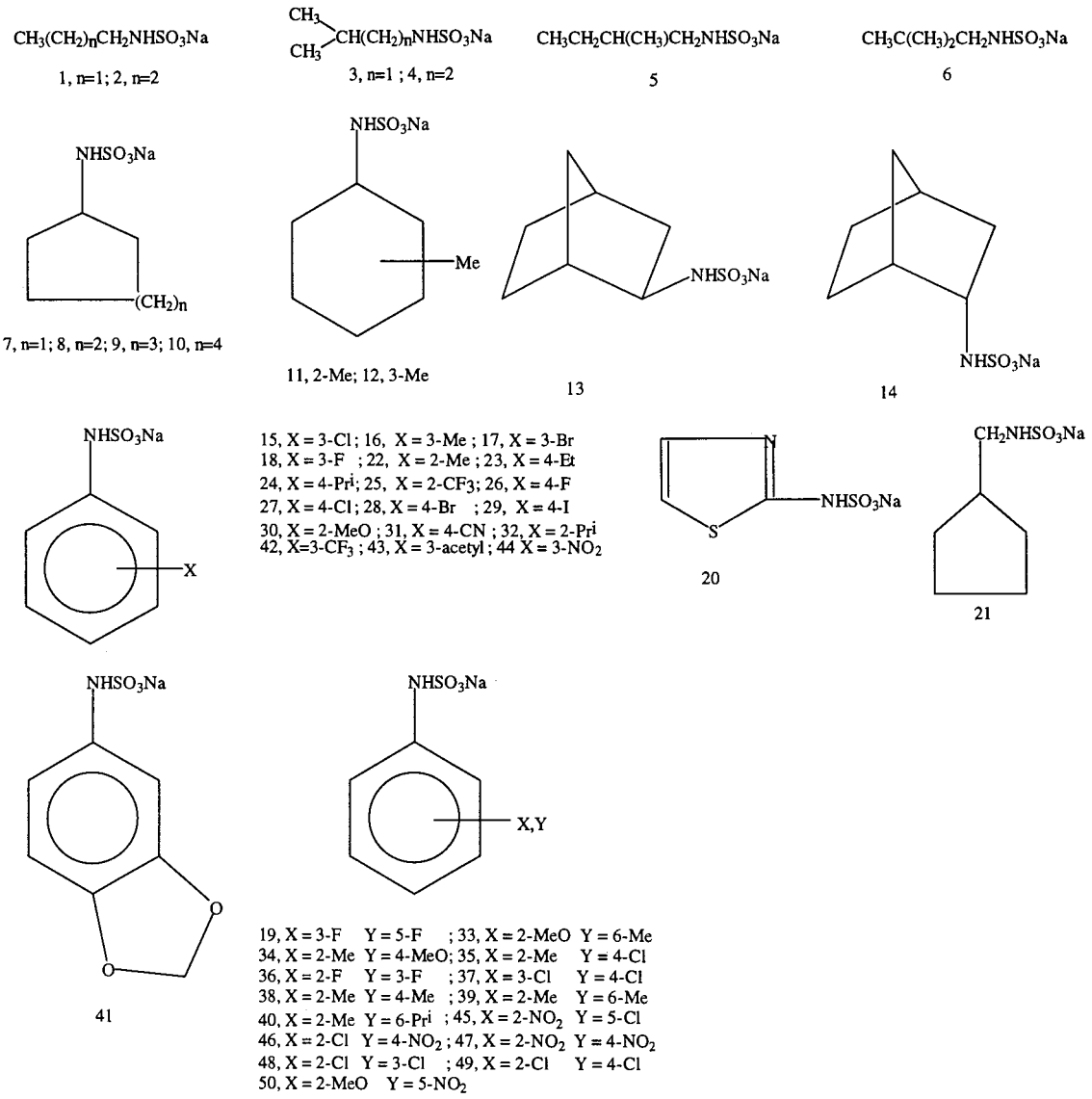

Figure 1. Fifty molecules used in this work. Molecules **1**−**21** taste sweet, **22**−**41**, sweet-bitter, and **42**−**50**, bitter.

initially formed ammonium or pyridinium salts were treated with sodium hydroxide, and after isolation, the crude sodium sulfamates were recrystallized several times from aqueous ethanol until they were free of sulfate (barium chloride test) and sodium chloride (silver nitrate test) and gave a quantitative precipitate of barium sulfate when their aqueous solution was heated with concentrated HCl in the presence of barium chloride. The sulfamates were characterized by IR and C, H, and N analysis.

Methodology. These sulfamates form a discrete set of molecules with similar shapes and properties and are therefore well suited to QSAR techniques. There is very little information about the sweet taste receptor, and therefore QSARs offer the best method for elucidating the important features of the molecules that give rise to the sweet taste.

Structures. The structures of all 50 molecules (21 sweet, 20 sweet-bitter, and 9 bitter) were built using CERIUS2 software (Molecular Simulations Inc., 1996). The sulfamate group was given an identical conformation in all of the compounds so that the only variation was in the R group. For all of the molecules, a charge of −1 was applied to the whole molecule using the PM3 Hamiltonian in the MOPAC 6.0 semiempirical software package (1995). In all cases, the positive counterion was not considered. For many of the sulfamates the lowest energy conformation was obtained straightforwardly by energy minimization, which was carried out using the default Dreiding2 force field. However, for some of the molecules, it was necessary to carry out a conformational analysis, which was done using the CERIUS2 software. In all cases only one conformation of the molecule was used in subsequent calculations. It can be argued that these lowest energy conformations may not necessarily be the active ones but as there is no evidence as to the active conformations, it seems logical to use the lowest energy calculations in this work.

Calculations of Properties. Values of relative sweetness were taken from the literature or obtained from a taste panel (see Experimental Procedures). Values of k_w were experimentally measured using reversed-phase HPLC (see Experimental Procedures). Corey−Pauling−Koltun volumes (V_{CPK}) were measured as described previously (Spillane et al., 1994). Third-order molecular connectivities ($^3\chi_m$) were calculated using the method of Kier and Hall (1976). A ring connectivity factor is not required in these calculations. A typical calculation (for compound **8**) is shown in

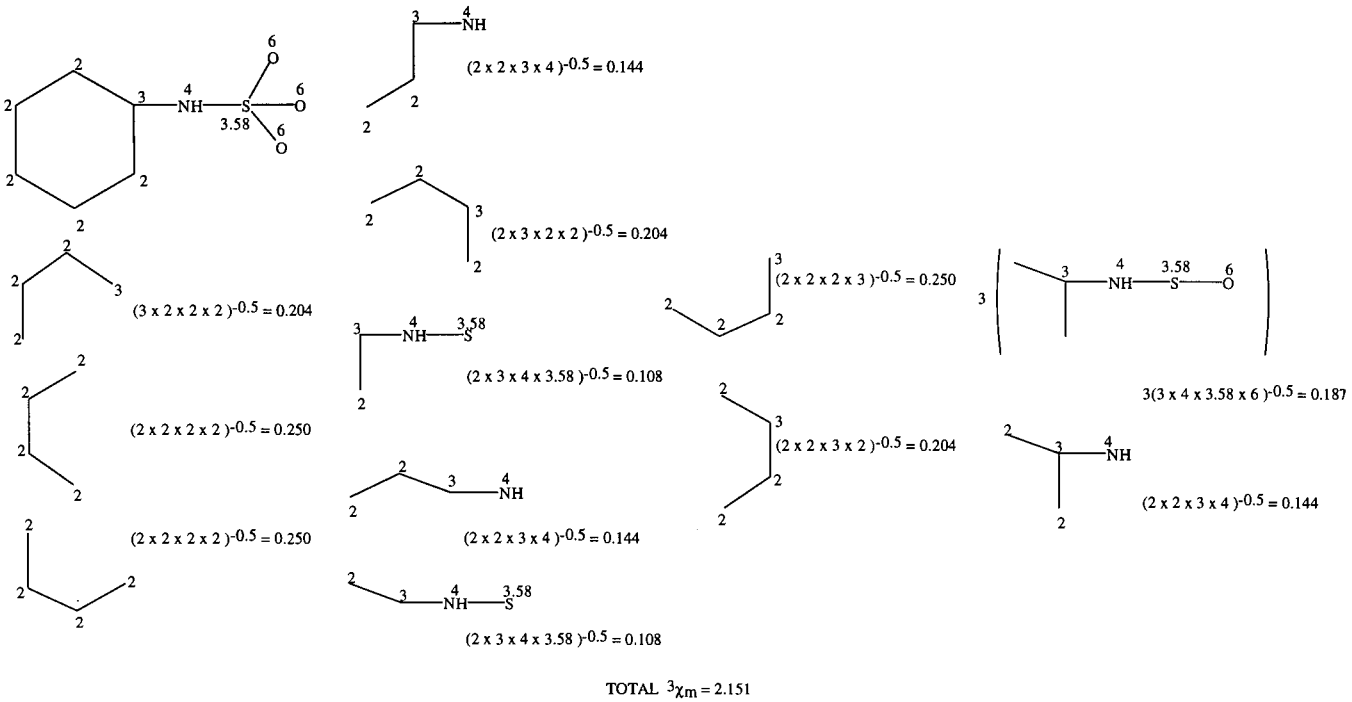

Figure 2. Method of calculation of the Kier $^3\chi_m$ parameter P2 for molecule **8**.

Table 1. RS of Sulfamates Compared to 3% (w/v) Sucrose Solution

no.[a]	range sampled, mol/L	concentration equivalent,[b] mol/L	RS[c]
4	0.015−0.0015	0.0233 ± 0.0066	6.8
5	0.05−0.0005	0.0081 ± 0.00094	19.6
6	0.06−0.0006	0.0099 ± 0.0021	16.0
11	0.04−0.0008	0.0073 ± 0.0014	15.8
12	0.05−0.0005	0.0032 ± 0.00085	42.9
13	0.008−0.0002	0.0018 ± 0.00037	70.5
14	0.05−0.002	0.0149 ± 0.0047	8.7
19	0.01−0.00025	0.0051 ± 0.00072	24.5
20	0.05−0.001	0.0165 ± 0.0043	8.6

[a] Eight tasters were used for each compound except **4**, for which five tasters were used. [b] Each entry is the pooled mean value ± standard deviation for all tasters. [c] RS is defined as the concentration of the standard sucrose solution, mg/mL i.e., 30/the concentration of the equivalent sulfamate solution, mg/mL.

Figure 2. Taft σ^* values were taken from the literature or calculated. The k_w, $^3\chi_m$, and σ^* values are included in Table 2 as parameters P1, P2, and P3, respectively. In addition to these values, a wide range of steric and electronic properties were calculated in the software packages TSAR (Oxford Molecular Ltd., 1996) and CERIUS2 (Molecular Simulations Inc., 1996). For each structure, one conformation was input and used by the software to calculate molecular properties such as χ_m and χ_v connectivity indices, the Verloop parameters of size and shape, the capacity factor log k_w, molecular surface areas and volumes, molar refractivity, and solubility. Capacity factors considered to be a measure of hydrophobicity were calculated from the database in Hansch et al. (1991). In addition, electronic properties such as dipole moments, HOMO and LUMO energies, heat of formation, and solubility were calculated using the MOPAC program interfaced to CERIUS2. Solubilities were calculated directly in CERIUS2. The full list of parameters used, P1, ..., P33, and their values are given in Table 2.

CLASSIFICATION ALGORITHMS

Discriminant Analysis (DA). DA is a supervised learning technique. It includes the dependent parameter and selects the parameters important for discrimination between different sets. This uses a Mahalanobis algorithm to select important parameters and uses this to plot the data in multidimensional space. For each class of molecules, the distance of each point from the center of each cluster is measured, and this score then represents the classification of the molecule. These can then be plotted in 2D to show the class membership.

Principal Component Analysis (PCA). PCA was carried out using TSAR software on various subsets of the parameters, using particularly those established by the genetic algorithm as the most important. It is important for the success of the method that the parameters included in the calculation are not highly correlated. The PCA allows the replacement of a large number of parameters by a much smaller number of principal components, which explain much of the discrepancies within the dataset. In PCA, the dependent parameter, in this case log(RS), is not included explicitly, and so all 50 compounds (the sweet, sweet-bitter, and bitter compounds) could be included. By plotting the principal components, it is often possible to cluster the data and highlight molecules in different families.

QUANTITATIVE/PREDICITIVE ALGORITHMS

Regression Analysis: Selection of Parameters. A regression equation relates a dependent parameter, in this case log(RS) to a set of parameters in a linear equation

$$\log(\mathrm{RS}) = a_0 + a_1p_1 + a_2p_2 + a_3p_3 + \ldots + a_np_n$$

where p_i are parameters and a_i constants calculated to get the best fit to the log(RS) value. In previous work (Spillane et al., 1996) we considered only four parameters, log(k_w), $^3\chi_m$, σ^*, and V_{CPK}, and established which

Table 2. Parameters Used in This Study[a]

ID	log(RS)	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P32	P33
1	−0.20	−0.66	0.61	−0.12	4830.00	152.40	108.80	−9.01	9.23	1.27	−3.98	−21.84	−11.89	2.85	18.93	126.50	22.99	86.10	89.73	0.91	0.78	−4.22	3.70	2.89	2.71	1.41	0.71	0.00	5.14	1.65	1.94	2.27	3.48	3.52
2	0.54	−0.24	0.89	−0.13	5343.30	172.00	126.60	−9.00	9.25	1.20	−3.78	−28.50	−11.88	3.38	34.18	189.40	27.32	128.80	136.10	0.57	1.06	−4.29	3.71	2.87	3.41	1.91	1.00	0.00	6.13	1.65	1.93	2.30	4.52	4.61
3	0.46	−0.27	1.15	−0.13	5343.30	170.30	126.40	−8.96	9.27	1.20	−3.64	−27.01	−11.85	3.25	34.24	150.80	31.37	102.30	106.20	0.57	1.32	−4.20	3.78	2.79	3.58	1.73	1.73	0.58	5.20	1.65	1.93	3.31	4.13	4.47
4	0.83	−0.07	1.49	−0.16	5856.60	195.70	144.50	−8.95	9.52	1.15	−3.44	−31.36	−11.82	3.78	38.75	228.50	33.06	156.40	163.30	1.10	1.66	−4.27	3.74	2.91	4.28	2.27	1.80	0.41	6.24	1.65	2.98	3.32	3.65	3.67
5	1.29	0.09	1.49	−0.16	5856.60	201.30	144.90	−8.97	9.46	1.15	−3.44	−30.69	−11.78	3.78	38.75	218.10	35.76	148.70	155.50	1.63	1.84	−4.14	3.75	3.21	4.28	2.27	1.80	0.41	6.47	1.65	2.30	3.33	4.04	4.45
6	1.21	0.06	2.41	−0.17	5856.60	196.60	144.40	−8.96	10.33	1.15	−3.33	−30.24	−11.75	3.86	37.41	217.60	36.20	151.40	152.00	1.24	2.58	−4.89	4.10	1.13	4.50	2.00	3.00	2.00	5.22	1.65	3.02	3.31	4.29	4.47
7	1.00	−0.21	1.90	−0.20	5715.80	173.40	132.90	−8.84	9.75	1.24	−4.62	−24.15	−11.59	3.44	35.52	142.00	38.85	96.37	96.77	1.45	2.07	−2.74	5.99	1.81	3.54	2.50	1.77	0.00	4.98	1.87	2.54	3.41	4.04	4.07
8	1.61	0.11	2.15	−0.15	6229.10	200.80	150.50	−8.83	10.77	1.18	−4.42	−32.05	−11.58	3.97	40.03	215.00	41.52	144.30	153.90	0.42	2.32	−4.51	4.47	1.27	4.24	3.00	2.12	0.00	4.59	1.89	2.69	4.20	4.33	4.90
9	1.53	0.40	2.40	−0.13	6742.30	216.30	168.60	−8.80	10.43	1.14	−4.21	−34.79	−11.53	4.50	44.55	266.90	48.07	175.80	195.00	1.22	2.57	−4.87	4.59	1.15	4.95	3.50	2.48	0.00	6.45	2.16	2.97	3.51	3.63	3.65
10	1.44	0.66	2.65	−0.13	7255.60	235.90	185.40	−8.79	10.82	1.11	−4.01	−38.34	−11.49	5.03	49.07	319.50	63.32	206.50	235.40	1.24	2.82	−3.97	4.21	3.09	5.66	4.00	2.83	0.00	6.82	2.36	2.94	3.86	3.97	4.15
11	1.20	0.43	2.85	−0.16	6742.30	227.30	167.70	−8.76	10.47	1.15	−4.11	−38.93	−11.48	4.34	44.61	236.70	58.92	148.40	174.70	1.26	3.02	−4.80	4.47	1.46	5.11	3.39	2.74	0.29	6.24	1.98	3.15	3.45	4.04	4.48
12	1.63	0.45	2.77	−0.16	6742.30	216.80	167.50	−8.76	10.71	1.15	−4.07	−42.07	−11.51	4.37	44.61	260.70	55.23	167.50	192.00	1.10	2.94	−2.86	5.83	1.12	5.11	3.39	2.74	0.29	6.25	1.98	3.43	3.91	4.08	4.50
13	1.85	0.20	3.20	−0.15	6601.60	197.10	156.50	−8.83	10.96	1.22	−5.11	−11.42	−11.60	3.90	41.44	261.40	40.23	181.10	184.20	0.40	3.78	−4.68	4.50	1.19	4.69	3.45	3.12	0.41	5.84	1.89	2.52	3.47	4.20	4.23
14	0.94	0.17	3.20	−0.15	6601.60	204.80	155.90	−8.82	11.11	1.22	−5.11	−10.11	−11.58	3.90	41.44	253.10	42.77	175.10	177.60	1.55	3.78	−4.67	4.51	0.98	4.69	3.45	3.12	0.41	5.10	1.90	2.83	3.20	4.52	4.78
15	1.42	0.22	1.68	0.85	8472.00	187.10	137.40	−8.12	15.33	1.50	−6.79	−2.01	−10.84	4.56	49.17	330.50	64.81	203.80	252.00	1.01	1.85	−4.46	4.48	2.23	4.58	3.46	3.19	1.21	6.31	1.71	2.05	3.23	4.41	4.43
16	1.19	0.11	1.58	0.48	7493.60	192.30	139.90	−7.87	15.48	1.33	−6.60	−7.99	−10.70	4.63	47.17	270.40	49.19	171.30	203.40	0.48	1.75	−4.79	3.98	2.30	4.39	3.37	3.02	1.11	6.44	1.73	2.21	3.22	4.10	4.15
17	1.05	0.33	2.05	0.86	6910.00	195.30	144.00	−8.10	15.30	1.74	−8.55	9.45	−10.83	4.20	49.78	441.10	80.09	273.30	336.80	0.48	1.75	−2.52	1.34	1.08	5.39	3.87	3.89	1.61	6.49	1.73	2.16	3.23	4.65	4.66
18	1.09	−0.09	1.00	0.82	7044.60	175.70	125.50	−8.21	15.39	1.52	−6.61	−40.48	−10.86	4.33	44.02	275.70	49.20	174.90	207.50	1.46	1.17	−4.39	4.35	2.21	3.77	3.06	2.48	0.80	6.30	1.67	1.71	3.23	3.84	3.84
19	1.39	0.07	0.84	1.64	7156.90	175.80	128.40	−8.47	13.68	1.62	−6.52	−81.21	−11.01	4.47	45.14	346.60	63.69	216.20	263.30	0.47	1.01	−3.98	3.90	2.03	4.07	3.11	2.66	0.93	6.30	1.66	1.68	3.84	4.02	4.02
20	0.94	−0.12	1.54	0.74	7209.10	146.60	109.70	−8.10	13.93	1.63	−10.39	25.26	−11.04	0.16	24.86	205.10	32.14	135.20	150.80	0.71	1.21	−5.23	4.44	2.06	3.47	2.93	2.55	0.57	5.79	1.74	1.97	2.75	3.30	3.30
21	1.18	0.32	1.54	−0.01	6229.10	197.50	150.30	−9.01	10.11	1.19	−4.42	−32.09	−11.82	3.97	56.06	313.40	31.34	215.20	225.60	0.75	2.47	−2.40	5.77	2.06	4.41	2.89	2.39	0.29	5.91	1.65	2.38	3.10	5.34	5.40
22	bitter/sweet molecules				7493.60	191.90	138.80	−7.88	14.62	1.34	−19.14	−0.04	−10.78	−5.06	50.03	237.90	47.17	152.00	176.90	0.69	1.83	−3.94	4.19	2.37	4.39	3.37	3.02	1.11	6.31	1.72	1.96	3.25	4.33	4.37	
23	values not obtained				8006.90	219.10	157.60	−7.73	16.57	1.27	−18.84	−14.82	−10.63	−4.49	54.54	395.20	41.14	268.90	286.70	1.65	2.05	−5.33	4.04	2.44	5.09	3.93	3.12	0.96	8.43	1.77	3.16	3.16	3.23	3.23	
24					8520.10	236.30	174.90	−7.71	16.82	1.23	−18.62	−18.57	−10.62	−4.09	59.12	485.70	43.83	337.40	346.60	0.70	2.53	−5.34	4.08	2.53	5.96	4.31	3.81	1.19	8.42	2.44	2.57	3.23	3.24	3.28	
25					7686.30	198.80	148.30	−8.35	14.61	1.62	−20.65	−144.10	−10.97	−4.46	48.62	366.40	103.90	210.90	281.00	1.88	1.32	−4.27	3.08	2.11	5.02	3.68	3.09	0.99	6.32	1.97	3.13	3.33	5.06	5.16	
26					7044.60	177.10	124.20	−8.01	14.48	1.53	−6.73	−43.93	−10.82	4.04	45.78	319.80	33.94	216.00	233.40	0.71	1.17	−4.46	3.96	2.29	3.77	3.06	2.48	0.80	7.14	1.65	1.70	3.15	3.22	3.22	
27					8472.00	189.70	137.20	−8.03	14.97	1.51	−6.79	−3.83	−10.80	4.56	50.26	416.90	35.03	284.40	302.80	1.63	1.89	−4.61	3.94	2.34	4.58	3.46	3.19	1.21	7.79	1.78	1.84	3.13	3.20	3.23	
28					6910.00	195.50	142.70	−8.06	15.14	1.76	−8.68	8.33	−10.79	3.91	51.54	605.40	35.97	417.60	436.90	1.63	1.78	−2.57	1.28	1.19	5.39	3.87	3.89	1.61	8.04	1.87	1.93	3.11	3.18	3.23	
29					6910.00	203.60	153.00	−8.08	15.12	1.95	−7.89	21.70	−10.80	4.99	42.13	780.70	36.54	541.40	561.30	1.63	1.78	−2.72	1.26	1.25	5.95	4.15	4.38	1.89	8.37	1.99	2.08	3.05	3.12	3.23	
30					7838.40	199.50	146.40	−7.72	14.64	1.38	−13.59	−28.73	−10.60	0.81	46.45	256.70	65.92	152.50	195.70	1.58	1.71	−4.42	3.42	3.09	4.80	3.48	2.71	0.81	6.32	1.66	2.91	3.23	5.26	5.35	
31					6910.00	188.40	132.70	−8.28	13.18	1.49	−17.68	36.95	−10.95	−4.09	41.04	390.70	35.10	265.70	284.30	1.62	2.05	−4.01	4.03	2.08	4.33	3.79	2.92	1.11	8.38	1.66	1.69	3.15	3.23	3.23	
32					8520.10	232.70	174.00	−7.95	14.50	1.23	−18.60	−2.26	−10.79	−4.08	59.12	296.70	97.36	167.90	224.50	0.66	2.50	−3.47	4.80	1.77	5.96	4.31	3.81	1.19	6.30	1.76	3.29	3.37	5.71	5.71	
33					8399.80	218.30	163.10	−7.66	14.49	1.33	−22.56	−35.83	−10.58	−4.88	58.62	287.20	89.98	157.20	222.90	1.40	2.18	−4.18	4.57	1.16	5.72	3.84	3.43	1.26	6.33	1.78	2.73	4.39	5.29	5.33	
34					8399.80	226.00	164.10	−7.57	16.14	1.32	−13.06	−39.24	−10.57	0.28	49.45	410.50	52.66	274.70	300.50	0.56	2.19	−6.45	4.04	6.87	5.72	3.84	3.43	1.26	8.53	1.72	2.07	3.25	4.34	4.37	
35					9033.40	206.70	154.50	−7.97	14.11	1.43	−19.12	−4.15	−10.82	−4.40	55.64	424.80	49.04	286.60	309.70	1.58	2.34	−3.76	4.13	2.40	5.51	3.83	3.91	1.65	7.78	1.72	1.98	3.25	4.33	4.37	
36					7156.90	173.80	128.40	−8.31	15.35	1.62	−6.87	−71.93	−10.94	4.02	46.90	287.20	57.93	179.50	216.60	1.86	0.96	−3.55	4.31	2.59	4.07	3.12	2.64	0.90	6.31	1.65	1.72	3.23	3.95	3.95	
37					10010.00	209.70	152.60	−8.17	15.09	1.58	−7.07	−4.32	−10.84	5.00	55.87	485.10	68.35	312.70	364.50	0.70	2.66	−4.51	4.43	2.40	5.70	3.94	3.90	1.55	7.79	1.80	1.81	3.23	4.46	4.46	
38					8055.00	214.20	157.10	−7.70	15.78	1.28	−19.05	−9.29	−10.65	−4.62	55.40	316.90	48.91	209.20	233.00	1.72	2.24	−4.10	4.24	2.91	5.31	3.73	3.74	1.56	7.38	1.73	2.06	3.26	4.33	4.37	
39					8055.00	205.60	157.40	−7.72	15.76	1.27	−19.05	−15.93	−10.63	−4.62	55.40	298.40	74.20	174.40	230.50	0.70	2.25	−3.10	5.62	3.02	5.31	3.73	3.73	1.55	6.34	1.93	1.97	4.39	4.40	4.40	
40					9081.50	250.70	190.90	−7.69	15.66	1.20	−18.58	−24.54	−10.62	−3.69	64.49	417.50	107.70	247.00	318.90	0.69	3.00	−4.91	4.15	2.95	6.89	4.68	4.52	1.63	6.98	2.53	3.62	4.33	4.92	4.96	
41					8090.60	194.90	141.90	−7.80	16.67	1.52	−18.81	−53.46	−10.60	−0.20	45.54	399.40	51.59	262.30	296.70	1.51	2.06	−10.63	2.84	4.63	4.83	3.73	3.06	0.90	7.40	1.76	2.11	3.22	3.64	3.65	
42	bitter molecules				7686.26	219.45	150.31	−8.25	15.13	1.60	−20.56	−141.89	−10.98	−4.39	48.62	443.25	89.85	276.97	334.20	1.60	1.27	−5.58	2.27	2.32	5.02	3.68	3.09	0.99	6.77	1.95	3.16	3.21	5.02	5.11	
43	values not observed				8461.00	213.52	153.70	−8.06	13.14	1.39	−7.39	−30.52	−10.92	2.64	52.90	376.30	78.36	227.85	289.04	1.69	2.02	−3.95	4.27	1.81	5.30	4.02	3.45	1.19	7.11	1.72	2.31	3.21	5.56	5.58	
44					8007.26	187.44	137.60	−8.37	15.31	1.58	4.38	48.97	−11.04	3.56	48.70	361.73	74.74	226.62	271.86	0.64	1.73	−10.11	3.23	4.60	4.65	3.82	3.21	1.18	6.33	1.95	3.13	3.29	4.37	4.39	
45					9547.04	203.75	152.95	−8.57	14.56	1.65	4.54	47.84	−11.09	4.32	54.31	433.86	160.40	212.06	342.84	0.35	2.26	−7.81	5.25	3.89	5.77	4.28	4.09	1.72	6.34	2.20	2.90	4.37	4.50	4.51	
46					9547.04	206.27	153.46	−8.51	12.42	1.64	4.36	46.35	−11.11	4.19	54.31	516.30	100.20	321.36	391.47	0.17	2.34	−9.71	1.14	1.14	4.77	4.28	4.09	1.72	7.54	2.38	2.77	3.38	4.08	4.57	
47					9082.34	223.16	153.20	−8.88	11.79	1.71	15.72	96.20	−11.30	3.32	53.83	533.16	117.10	328.69	403.12	4.00	2.15	−4.37	1.19	1.08	5.84	4.64	4.11	1.69	7.55	2.38	2.81	3.93	4.44	4.50	
48					10011.74	198.92	152.73	−8.16	14.19	1.58	−6.96	−1.09	−10.92	5.08	54.78	450.94	72.03	285.82	341.27	0.14	2.73	−4.45	3.48	2.63	5.70	3.94	3.90	1.55	6.34	1.71	1.84	3.17	4.57	4.57	
49					10011.74	198.65	153.15	−8.08	14.77	1.57	−6.81	−3.35	−10.86	5.19	54.78	476.32	92.38	290.33	366.14	0.71	2.45	−4.74	3.79	2.75	5.70	3.93	4.08	1.75	7.81	1.79	2.17	3.15	4.55	4.58	
50					8913.48	218.91	162.80	−8.04	14.84	1.52	−1.97	9.77	−10.87	2.60	46.56	523.76	100.92	326.44	396.96	4.00	2.11	−10.06	1.28	1.12	5.98	4.30	3.62	1.33	7.58	2.34	2.74	3.80	4.84	5.48	

[a] Activity is log(RS). Log of RS measured using a taste panel. Details in text. Molecule 20 was estimated using literature data (Nofre and Pautet, 1975a). P1: log k_w, measured using HPLC. The value for molecule 20 was estimated by taking an average of the log k_w values predicted by the following equations found for the sulfamates, e.g., $V_{CPK} = 115.419 \log(k_w) + 166.069$, $r^2 = 0.716$, $n = 19$; $^3\chi_m = 1.909 \log(k_w) + 1.735$, $r^2 = 0.527$, $n = 19$. P2: $^3\chi_m$, molecular connectivity index. Calculated as described in the text and illustrated in Figure 2. P3: Taft σ^* values taken from the literature, but values for molecules **13** and **14** were estimated using σ_p^+ values (Hansch et al., 1991). The value for molecule 20 was estimated from σ_1 values (Taylor and Wait, 1986). P4: sum of atomic polarizabilities (CERIUS2). P5: molecular surface area/Å^2 (CERIUS2). P6: molecular volume/Å^3 (CERIUS2). P7: energy of LUMO/ev (MOPAC). P8: dipole/D (MOPAC). P9: density/g cm^{-3} (CERIUS2). P10: free energy of dissolvation in water/kcal mol^{-1} (CERIUS2). P11: heat of formation/kcal mol^{-1} (MOPAC). P12: energy of HOMO/ev (MOPAC). P13: log of partition coefficient (CERIUS2). P14: molar refractivity/cm^{-3} (CERIUS2). P15: principal moment of inertia/10^{-39} g cm^{-2} (CERIUS2). P16: principal moment of inertia/X component 10^{-39} g cm^{-2} (CERIUS2). P17: principal moment of inertia/Y component 10^{-39} g cm^{-2} (CERIUS2). P18: principal moment of inertia/Z component 10^{-39} g cm^{-2} (CERIUS2). P19: superdelocalizability (CERIUS2). P20: $^3\chi_v$ molecular connectivity index (CERIUS2). P21: X component of dipole moment/D (CERIUS2). P22: Y component of dipole moment/D (CERIUS2). P23: Z component of dipole moment/D (CERIUS2). P24: $^0\chi_v$ molecular connectivity index (TSAR). P25: $^1\chi_v$ molecular connectivity index (TSAR). P26: $^2\chi_v$ molecular connectivity index (TSAR). P27: $^3\chi_v$ molecular connectivity index (TSAR). P28: Verloop box L parameter (TSAR). P29: Verloop box B1 parameter (TSAR). P30: Verloop box B2 parameter (TSAR). P31: Verloop box B3 parameter (TSAR). P32: Verloop box B4 parameter (TSAR). P33: Verloop box B5 parameter (TSAR).

combination gave the best fit. However, in the present work we calculated 30 additional parameters for each molecule, although many of these properties were not independent and therefore could not be used simultaneously in a regression analysis. It was clearly necessary to reduce the number of parameters significantly, taking into account that there were only 21 molecules with known RS values (the RS value for the sweet-bitter compounds being undetermined). One method for doing this involves inspection of the correlation matrix. Thus, when two parameters were correlated with $r^2 > 0.8$, then the parameter with the lowest correlation with log(RS) was removed. Gradually the number of parameters in the regression equation can be reduced until an acceptable number is reached. The optimum number of parameters for a given number of molecules is difficult to estimate, but it is well-known that the dangers of chance correlation increase significantly with larger numbers of parameters (Topliss and Edwards, 1979).

Genetic Algorithms. The genetic algorithm method can also be used to establish the most suitable parameters for a regression equation (Rogers and Hopfinger, 1994). The method is applied as follows. All parameters, however correlated, can be included. Many regression equations are developed using a random choice of parameters. The actual number of parameters or indeed the form of the equation to be used can be selected by the user or randomly. The genetic crossover operator combines random parts of these equations to give two new equations. If either of these are improvements on these previous equations, as measured by the lack of fit (LOF) (Friedman, 1988), then they are adopted as replacements. The smoothing parameter *d* in this equation was kept at the default 1.0 value. The effect of changing *d* was investigated but was found to be quite small. This procedure is repeated as many times as desired, although CERIUS2 has a limit of 10^6. The end result of the genetic algorithm process is a set of regression equations using a limited number of the available parameters. From the regression equations chosen, it is possible to observe those parameters that appear most often and therefore be able to see which parameters are most closely related to sweetness. The genetic algorithm calculation was carried out using the program described previously interfaced with the CERIUS2 software. The genetic algorithm method also allows the inclusion of nonlinear methods of mapping data such as spline, quadratic, and offset quadratic and may include any combination of these within an equation. This is in contrast with the usual multiple regression method, which is restricted to linear terms.

Neural Networks. Neural networks are adaptive learning algorithms based on the brain structure and have been applied in QSAR for some time (Andrea and Kalayeh, 1991; Aoyama et al., 1990; So and Karplus, 1996). The neural network contains a number of input nodes, the independent parameter set, and usually one output node, the dependent parameter, in this case log(RS). Between there are a number of hidden nodes in one or more layers. Each node feeds a value forward to the next node. This value is assigned a weight. Each node sums the input weights and, if they are greater than a threshold, transmits a value to the next node. Two sets of data are needed: one to train the network and another to test it. The network is trained by using numerical or Monte Carlo techniques to adjust the weights on each node until the overall RMS errors are minimized. A major problem is that the number of hidden nodes cannot be accurately predetermined but must be established by trial and error. These calculations were carried out using the TSAR software. It is important not to include too many nodes in the network as this will lead to overfitting and memorizing the dataset, or too few nodes, which will lead to low predictivity.

Molecular Field Analysis. Molecular field analysis was carried out within the CERIUS2 software. The molecules were aligned on the $C-N-SO_3$ moieties using the lowest energy conformations. Potentials were sampled for a grid of points in space around the molecules with OH^-, H^+, and CH_3 probes. The points were 2 Å apart within a grid of $10 \times 12 \times 12$ Å. Other methods of alignment and other grids did not prove as successful. Indeed, using a grid of $20 \times 20 \times 18$ Å gave significantly worse results. The potentials for each probe provided $6 \times 7 \times 7$ values per molecule, making 882 in all. These values were then subjected to a genetic algorithm to create a suitable linear regression algorithm with log(RS). Although a methyl probe was used, the resulting grid points were not selected by the genetic algorithm; indeed, five H^+ and one OH^- sites were selected.

RESULTS AND DISCUSSION

PCA and DA. DA was able to separate the 50 molecules into their 3 classes using 11 components. The parameters used were P4 (sum of atomic polarizabilities), P21 (*X* component of dipole moment), P7 (energy of LUMO), P13 (log of partition coefficient), P18 (principal moment of intertia, *Z* component), P27 ($^3\chi_v$), P32 (Verloop B4), P12 (energy of HOMO), P15 (principal moment of intertia), P24 (molecular connectivity index $^0\chi_v$), and P30 (Verloop box B2).

The results obtained using this technique are shown in Figure 3. Several tests were then carried out in which the activity of each molecule was assigned randomly. Though some clustering was subsequently observed, no class separation was ever achieved.

The PCA was carried out on all 50 molecules, those that tasted sweet as well as those that tasted sweet-bitter and bitter. PCA requires independent parameters as input, and we used the genetic algorithm technique along with a correlation matrix to choose those parameters to be input to this analysis. However, even by using all of the parameters in the dataset, the PCA technique was unable to give separation of all three classes. For sweet and sweet-bitter molecules the technique fared better, being clearly able to distinguish between the classes. The results for this are listed in Table 4. The PCA on parameters P4, P6, P7, P10, P19, P20, P23, P28, and P31 gave rise to principal components that explained, respectively, 0.385, 0.622, 0.744, 0.836, 0.904, 0.945, 0.974, 0.987, and 1.000 of the data. PC1 was plotted against PC2 to obtain the graph shown in Figure 4. We then repeated the PCA calculation for the sweet-tasting molecules only. However, neither a 2D nor a 3D graph incorporating the first two or first three PCs gave rise to any clustering. The clustering method then only gives the distinction between classes and is not able to distinguish between degrees of sweetness.

Regression Analyses. RS data have been reported previously by us for compounds **1**−**3**, **7**−**10** (Benson and Spillane, 1976), and **15**−**18** (Spillane et al., 1989). Eight

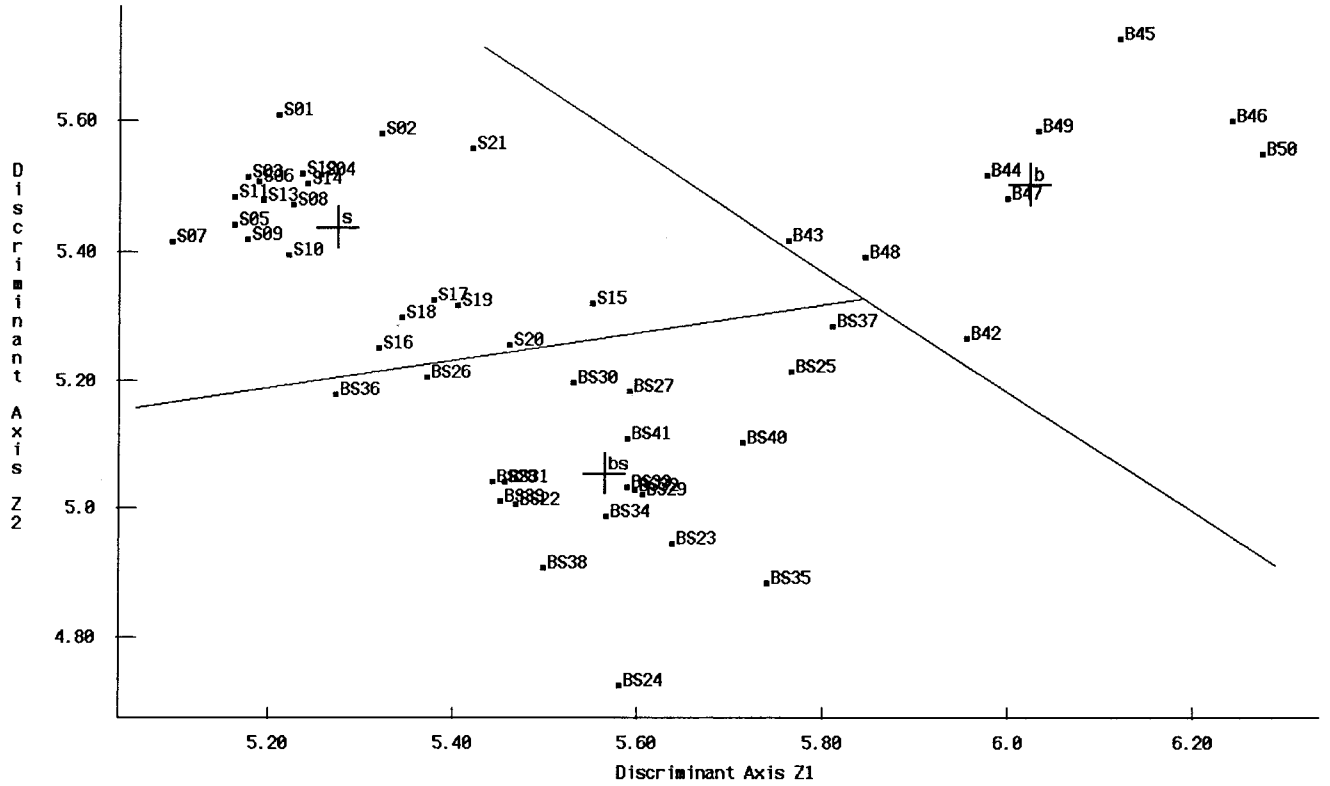

Figure 3. Results from the DA. Molecules with sweet taste (**1**−**21**) are designated S, molecules with sweet-bitter taste (**22**−**41**), BS, and molecules with bitter taste (**42**−**50**), B.

Table 3. Results from Genetic Algorithm Calculations[a]

				predictions for log(RS) for molecules **3**, **14**, and **18**[b]		
type of eq	LOF	r^2	eq	**3** (0.46)	**14** (0.94)	**18** (1.05)
linear	0.125	0.624	−1.34 + 0.29(P26) + 0.01(P6)	0.77	1.74	1.27
linear	0.126	0.624	−0.25 + 0.49(P25) − 0.52(P21)	0.58	1.25	1.36
linear	0.112	0.754	−1.22 + 0.37(P2) + 0.47(P28) − 0.24(P32)	0.49	1.11	1.07
linear	0.113	0.752	0.07 + 0.29(P26) + 0.01(P6) − 0.27(P32)	0.77	1.46	1.19
linear	0.152	0.768	1.97 + 0.15(P8) − 0.33(P32) + 0.79(P1) + 0.01(P16)	0.59	1.21	1.21
linear	0.163	0.751	0.84 + 0.01(P6) − 0.43(P30) − 0.44(P6) + 0.02(P14)	0.89	1.28	1.19
spline[c]	0.178	0.784	1.19 + 0.0001(⟨P4−5487⟩) − 0.596(⟨P9−1.1⟩) −2.370(⟨2.848-P31⟩)	0.88	2.72	1.78
quadratic	0.116	0.745	1.07 + 0.042($P25^2$) + 0.003($P14^2$) − 0.0498($P32^2$)	0.73	1.27	1.40
quadratic	0.113	0.827	−0.51 + 0.034($P31^2$) + 0.633($P27^2$) − 0.032($P33^2$)	0.73	1.27	1.40
free	0.110	0.853	0.089 + 0.239(P24) − 0.8662(⟨2.483-P25⟩) + 0.898(⟨0.139-P21⟩)	0.56	0.47	1.91

[a] For parameter numbers P*n* refer to Table 2. [b] Experimental values in parentheses. [c] Truncated power splines such that $\langle f(x)-a\rangle = 0$ if $(f(x)-a) < 0$ else $= f(x)-a$.

Table 4. Results of the PCA

	principal component								
parameter	1	2	3	4	5	6	7	8	9
P4	0.484	−0.023	−0.066	−0.094	−0.219	0.115	0.655	−0.331	0.391
P6	0.329	0.461	0.039	0.310	0.070	−0.007	−0.307	0.635	−0.281
P7	0.459	−0.235	−0.108	−0.269	−0.133	0.016	0.110	0.199	−0.762
P10	−0.435	0.134	−0.028	0.351	−0.127	0.607	0.411	−0.029	−0.337
P19	0.084	−0.141	0.861	0.032	0.423	0.011	0.211	−0.009	−0.085
P20	0.166	0.557	0.107	0.348	−0.191	−0.382	0.239	0.538	−0.009
P23	0.172	−0.333	−0.414	0.550	0.587	−0.111	0.150	0.080	0.008
P28	0.383	−0.251	0.185	0.374	−0.321	0.471	−0.409	0.268	0.227
P31	0.214	0.459	−0.153	−0.366	0.500	0.487	−0.062	0.280	0.126
fraction of variance explained	0.385	0.237	0.121	0.092	0.068	0.041	0.029	0.015	0.011
cumulative total of variance explained	0.385	0.622	0.744	0.836	0.904	0.945	0.974	0.989	1.000
eigenvalue	3.463	2.137	1.093	0.831	0.612	0.367	0.262	0.132	0.103

more compounds, **4**−**6**, **11**−**14**, and **20**, have been synthesized in this work, and their RS values are given in Table 1. An RS value for compound **21** has been derived from literature data (Nofre and Pautet, 1975a,b). In Table 2 the log(RS) values for all 21 compounds are given. We first continued our regression analysis methodology established previously (Spillane et al., 1996) developing equations involving just four parameters, the logs of the capacity factors (k_w), Taft sigma star (σ^*) values, Corey−Pauling−Koltun volumes (V_{CPK}) and third-order molecular connectivities ($^3\chi_m$). Third-order molecular connectivities have proved useful in

Table 5. Results of the DA[a]

parameter	constant	parameter	constant
P6	−0.005	P25	−0.585
P8	0.593	P28	0.067
P27	0.185	P31	0.514
P21	−0.035	P32	0.030

[a] Composition of the discriminant axis in terms of the parameters used in the model.

Table 6. Results from Neural Network Calculations

network configuration	811	821	831	841	8421	881
training set RMS error	0.253	0.036	0.030	0.028	0.028	0.030
test set RMS error	0.262	0.322	0.345	0.309	0.339	0.262

	log(RS)						
molecule	exptl	calcd					
1	−0.222	−0.152	−0.224	−0.230	−0.215	−0.225	−0.233
3[a]	0.462	0.416	0.746	0.882	0.488	0.848	0.379
2	0.544	0.453	0.570	0.553	0.551	0.541	0.551
4	0.832	0.908	0.839	0.831	0.833	0.853	0.845
20	0.934	0.927	0.944	0.939	0.927	0.926	0.927
14[a]	0.939	1.515	1.616	1.603	1.601	1.600	1.601
7	1.000	0.992	0.988	0.999	1.000	1.002	1.005
17	1.049	1.048	1.002	1.040	1.039	1.035	1.054
18[a]	1.093	1.261	1.181	1.196	1.349	1.227	1.272
21	1.180	1.507	1.184	1.194	1.185	1.186	1.184
16	1.185	1.366	1.179	1.205	1.215	1.221	1.194
11	1.199	1.375	1.228	1.220	1.195	1.207	1.189
6	1.204	1.212	1.226	1.209	1.206	1.198	1.202
5	1.292	1.333	1.270	1.271	1.260	1.256	1.273
19	1.389	1.283	1.403	1.393	1.378	1.400	1.408
15	1.417	1.342	1.388	1.389	1.408	1.410	1.414
10	1.444	1.536	1.459	1.468	1.466	1.425	1.446
9	1.533	1.326	1.539	1.517	1.521	1.528	1.558
8	1.613	1.513	1.627	1.608	1.611	1.624	1.620
12	1.632	1.472	1.615	1.605	1.619	1.628	1.623
13	1.842	1.643	1.837	1.818	1.811	1.842	1.839

[a] Molecules in the test set.

correlating sweetness and structure (Daniel and Whistler, 1982) and were therefore preferred to lower order molecular connectivities. However, because of the interrelationship between log(k_w)and $^3\chi_m$ (see footnote, Table 2), they could not be used in the same regression equation. Log(k_w) values also correlate with V_{CPK} values and were not used in the same equation. As k_w was experimentally determined, it was decided to use it preferentially in this study.

Among the sweet-bitter group of compounds, **22**−**32** and **36**−**41** have been synthesized and assessed for taste quality (sweet, bitter, sour, salty) previously (Spillane et al., 1993, 1994). Compounds **33**−**35** have been synthesized in this work, and their taste assessment places them in the sweet-bitter category (see Experimental Procedures). Some years ago Greenberg (1990), using sweet threshold data for the seven sulfamates **1**−**3** and **7**−**10** for which such data were then available (Benson and Spillane, 1976), derived the equation

$$\log(1/c) = 0.68 \log(P) + 0.05$$
$$[n = 7, r^2 = 0.74, s = 0.33]$$

where c is the threshold concentration for sweetness and P is a hydrophobicity constant calculated from fragmentation constants. This is not a particularly good correlation and is very limited, with just seven data points giving an RS spread (least to most sweet) of only 65. However, it does point to the importance of hydrophobicity in the attempt to correlate sweetness quantitatively (Daniel, 1989). All available RS data on sulfamates have now been brought together (Table 2). The spread of RS is now >100-fold, and the diversity of structural types is evident. Capacity factors log k_w, which are considered to be a measure of hydrophobicity, were measured by reversed-phase HPLC. These values correlate very well with partition coefficients (Braumann, 1986) and the Hansch π parameter (Hansch et al., 1991) and for the highly water-soluble sodium sulfamates in this work are much more readily measured than partition coefficients in *n*-octanol/water. The following equation was derived:

$$\log(RS) = 1.21 \log k_W + 1.01$$
$$[n = 21, r^2 = 0.62, s = 0.29]$$

Because electronic factors feature prominently in several sweetness QSARs, Taft σ^* values were also used in the regression analysis, leading to the equation

$$\log(RS) = 1.12 \log k_w + 0.096\sigma^* + 0.995$$
$$[n = 21, r^2 = 0.63, s = 0.295]$$

These relationships are quite good considering the diversity of the structures (aliphatic, alicyclic, aromatic, and heterocyclic) encompassed in the analysis. Values of log k_w, $^3\chi_m$, and σ^* are shown in Table 2 as parameters P1, P2, and P3, respectively.

We then used the genetic algorithm method with all of the parameters to investigate whether preferable equations could be found and to establish the most appropriate parameters for further study. Input for the genetic algorithm does not preclude highly correlated parameters, and all values in Table 2 were permitted to be included.

In the program it is possible to select the equation required, and we chose linear regression with two, three, or four variables; spline with three or four variables; quadratic with three or four variables; and a "free method" in which the program selects the most useful form of each parameter in the equation. The best equations found, as indicated by the LOF and r^2 values and cross-validated r^2, are presented in Table 3. Whereas higher values of r^2 are desirable, the value of LOF should be at a minimum to give the best fit to the data. With linear equations, and two parameters, it is remarkable that the highest r^2 values are only 0.750, only slightly better than those obtained by us previously using just parameters P1−P4. It would seem therefore that this is the best value that can be obtained with two such parameters. With three parameters, the r^2 values increase signficantly, and many equations have r^2 values of ∼0.85. The LOF values can be compared directly with those for two parameters and indicate that fits with three parameters are significantly better. The quadratic fits show an increase in r^2 but also an increase in the LOF, which suggests that they do not represent a significant improvement. The spline method leads to increases in both r^2 and LOF. The best fit of all as measured by both the LOF (0.042) and the r^2 value (0.942) is found in the "free" calculation, when the program selects the most appropriate type of equation from those available and in this case chooses a combination of spline and linear with three parameters.

Whereas the r^2 value gives an idea of how well the equation fits the dataset, it is not a good measure of predictability. For this the cross-validated r^2 (XVr^2) is used. This is calculated by alternatively holding each

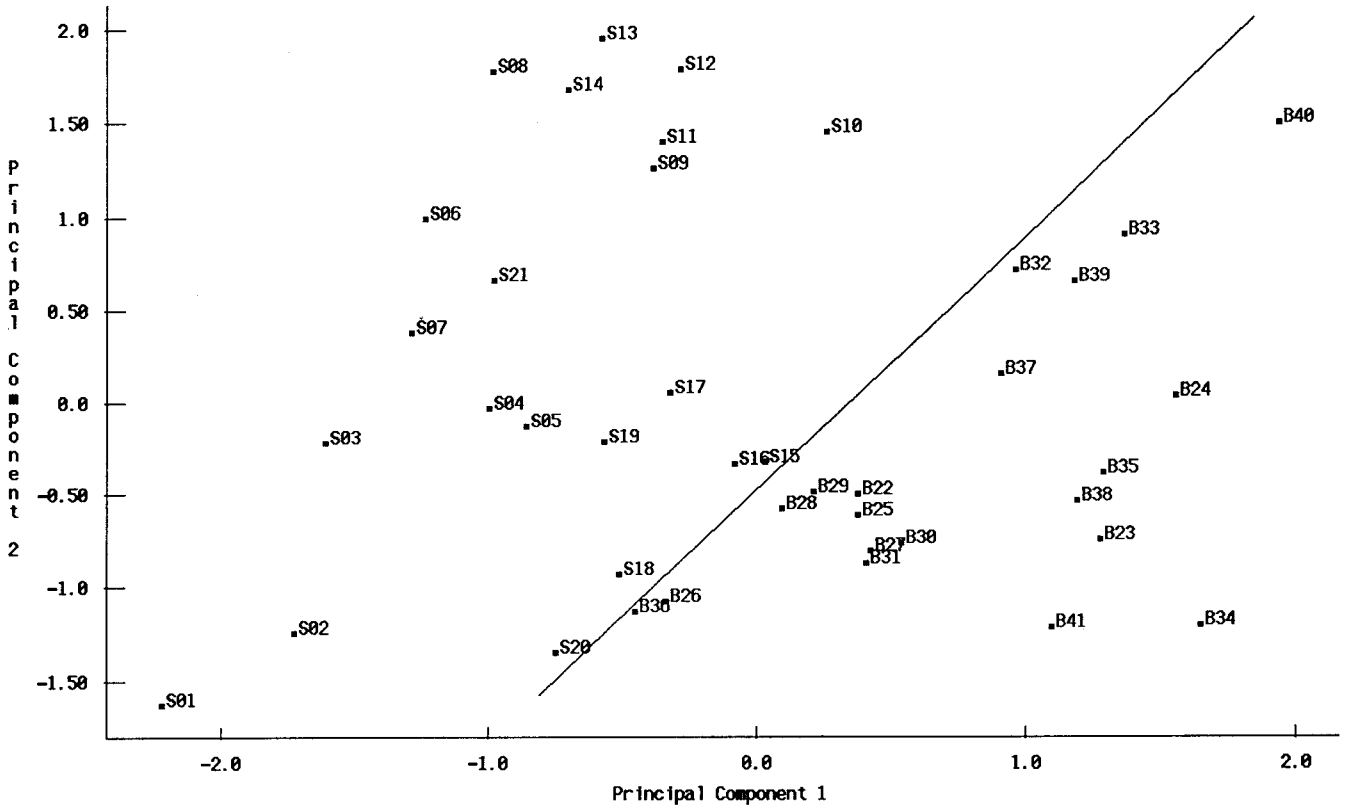

Figure 4. Results from the PCA. Plot of PC1 against PC2 shows the clustering between sweet and sweet-bitter compounds. Sweet molecules are numbered **1**−**21** and sweet-bitter molecules, **22**−**41**.

molecule out of the equation and comparing its least-squares error to that of all the others. For some of the equations, especially the free method ones, we can see that we are achieving high predictability. As a further test of the predictive power of the method, we omitted three compounds (**3**, **14**, and **18**) from the calculation. These were chosen to give the best range of sweetness value and structural type. In Table 3 are quoted the values predicted by each equation. Generally the predictive power is high for the best equations. To confirm the validity of the results, molecules were assigned random activities by two methods. First, the real activities of the molecules were randomly redistributed to the dataset, and second, a program was used to generate a random number in the range of −0.2 to 1.5, which was used as the log(RS) score for each molecule. The genetic algorithm technique was applied to both of these datasets using equation types previously having been found to give good results: linear with three and four parameters, spline with three parameters, and quadratic with four parameters. r^2 values of ∼0.55 with LOF scores of ∼0.2 were recorded. Cross-validated r^2 scores never rose above 0.4. The paramaters used by these equations were not in the set we had previously established as being important for sweetness. This leads us to conclude that we are observing a real trend in the data and not an artifact of the calculations.

Neural Networks. In the neural network, we used the same set of nine independent parameters that were used successfully in the PCA. Of the 21 molecules to be considered, 18 were taken as the training set and 3 as the test set. The test set comprised molecules **3**, **14**, and **18**. There were nine input parameters, and we varied the number of hidden nodes but found very little difference in the results as indicated by the RMS values for the training set and the test set (Figure 5). Results for the 9−3−1 network are perhaps the best with RMS values for the training set of 0.0157 and for the test set of 0.2326. The agreement between measured and calculated values for log(RS) are for molecule **3** (0.462) 0.421, molecule **14** (0.941) 1.436, and molecule **18** (1.389) 1.276.

Molecular Field Analysis. The molecular field analysis was carried out on the 21 sweet-tasting molecules, which were superimposed via the C−NH−SO_3 moiety. Genetic algorithm methods were employed to obtain regression equations using the potential values on the grid points. The best equation used six such points, five from the H^+ probe and one from the OH^- probe. The r^2 value was 0.958, LOF 0.045 and XVr^2 0.902. Although these were the best values found, many similar values were obtained using the genetic algorithm. The locations of these six points together with their coefficients in the regression equation are shown in Figure 6.

CONCLUSIONS

We have extended our set of sweetness data by introducing several new compounds, and we have used a variety of statistical techniques to analyze these data. We have used both DA and PCA to cluster the molecules into their classes, DA being able to separate sweet/sweet-bitter/bitter and PCA being able to separate sweet/sweet-bitter. The clustering of sweet and sweet-bitter compounds via the PCA was a particularly significant result as the PCA is an unsupervised learning method that requires no prior knowledge of the classification. The genetic algorithm approach has proved particularly useful not only in formulating good regression equations but also in establishing the most important structural parameters for subsequent use in

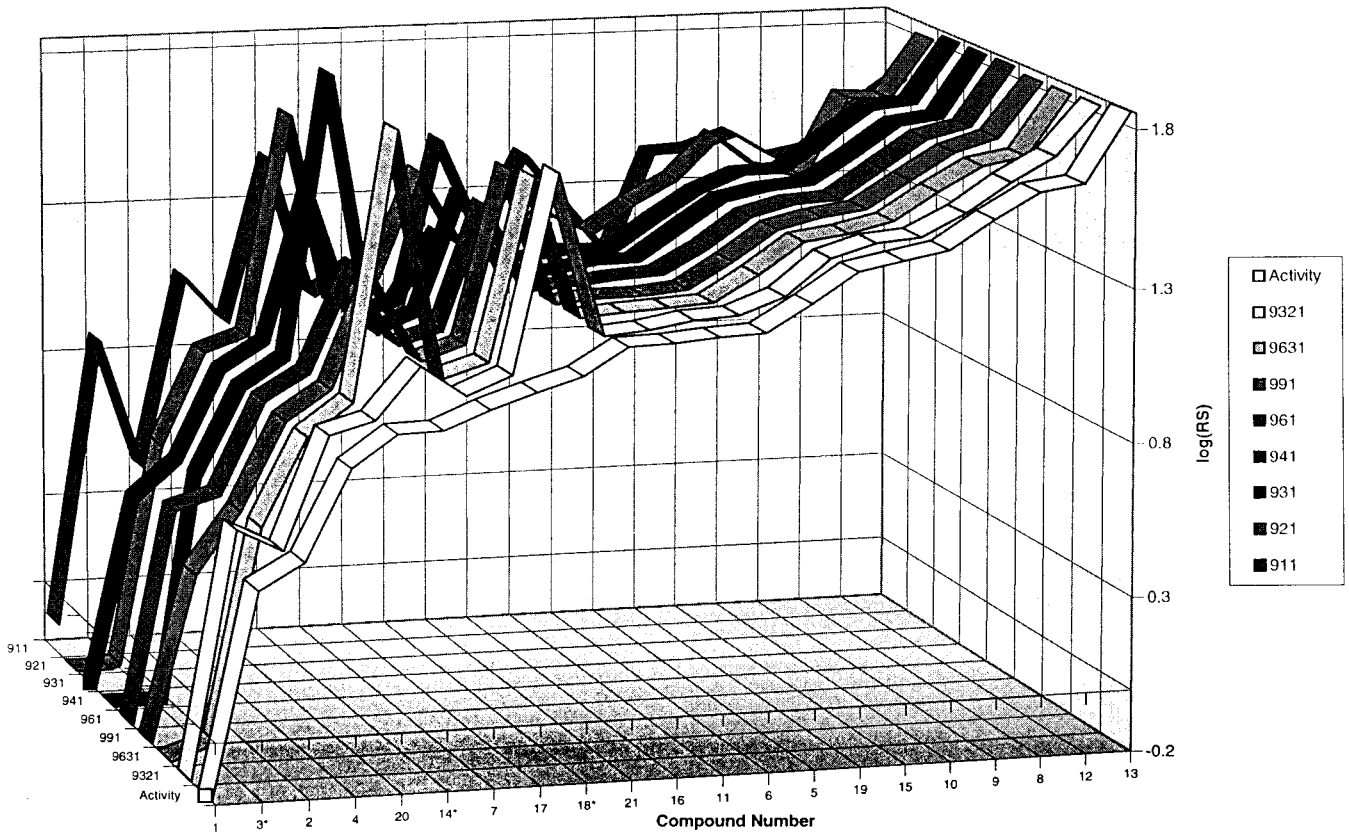

Figure 5. Results from the neural network analysis described in a three-dimensional plot. The *x*-axis represents the compound number. Compounds **3**, **14**, and **18** (marked with an asterisk) constitutue the test set. The *y*-axis represents the neural network configuration as described by the number of input, hidden, and output nodes respectively). The *z*-axis represents the log(RS) value obtained for each neural network (all shaded), and these are compared to the experimental values (unshaded)

QSARs. The established neural network technique proved to have some predictive power but generally was not so successful. However, the most valuable technique for this set of compounds proved to be the molecular field analysis, which provided the highest r^2 and XVr^2 ($r^2 = 0.958$, XV$r^2 = 0.902$) of all the techniques and is likely to have the best predictive power.

We now have developed several techniques that used together should allow us to classify the taste (sweet versus sweet-bitter versus bitter) for any new sulfamate and also give a confident prediction of the relative sweetness of the molecule. There were two major handicaps in our analysis. It would have been preferable in this analysis to have recourse to sweetness data on many more sulfamates; it also would have been preferable to have a larger range of activity data than that found for log(RS) of −0.20 to 1.85. These methods have been applied to a range of related compounds, the sulfamates. It remains to be seen whether they could be applied successfully to the whole range of molecules that have interesting taste properties.

EXPERIMENTAL PROCEDURES

Synthesis. Thirteen compounds were synthesized in this work, compounds **4**−**6**, **11**−**14**, **20**, **33**−**35**, **45**, and **46**. All gave satisfactory C, H, and N (within ±0.5 of the calculated values) except the following: **4** ($C_5H_{12}NSO_3Na{\cdot}0.5H_2O$) C, 30.9, requires 30.3; **11** ($C_7H_{14}NSO_3Na{\cdot}2.5H_2O$) C, 32.9, requires 32.3; and **35** ($C_7H_7NSO_3Na{\cdot}1.5H_2O$) H, 3.07, requires 3.69. Most of the sulfamates contain water of crystallization, which is normal for these salts (Benson and Spillane, 1976). Drying at high temperatures must be avoided since in the case of aromatic sulfamates, isomerization of the sulfamates can occur (Alexander, 1948). IR spectra were measured in Nujol mulls on a 983G Perkin-Elmer spectrophotometer, and all gave the usual characteristic bands associated with the sulfamate function (Vaugnat and Wagner, 1957; Nofre and Pautet, 1975a,b): 3400−3190 (ν NH), 1238−1210 (ν asym SO_3) 1203−1170 (ν-symm SO_3), 1071−1040 (ν sym SO_3), and 730−660 cm^{-1} (ν NS).

Sensory Analysis. In our laboratory we have previously determined RS values for compounds **1**−**3** and **7**−**10** (Benson and Spillane, 1976) and compounds **15**−**18** (Spillane et al., 1989). In the current work we have determined RS values for compounds **4**−**6**, **11**−**14**, **19**, and **20**. All RS values were determined compared to a 3% w/w sucrose solution as previously described (Spillane et al., 1989). The precision of the RS value for the "parent" sulfamate (cyclamate), **8**, is 41 ± 1.5, on the basis of four determinations by three different taste panels over a number of years. Thus, the deviation of this value is <4%. **11** has been determined previously as 15 (Daniel, 1989) in agreement with our value of 15. Generally, the RS values should be accurate to within 5%. Because of the very low RS of **1** and the estimations involved, the value obtained for this may be less precise. Similarly for **21**, an RS has been estimated (see footnote, Table 2) and may not be as accurate as the experimentally determined values.

The taste panel procedure for assessing the taste quality of **33**−**35** has been described in detail in earlier work (Spillane et al., 1994). In the present work six trained assessors were used, and the following are the percentages of assessors giving the tastes sweet, sour, bitter, salty, tasteless, and sweet aftertaste: **33**, 0, 0, 100, 0, 0, 100%; **34**, 0, 0, 100, 0, 0, 60%; and **35**, 0, 0, 100, 0, 0, 80%.

Capacity Factor (k_w) Measurements. k_w values were determined by reversed-phase HPLC using a Technopak 10C18 column. A Milton Roy CM4000 multiple solvent delivery system and an Altec refractive index detector were used in the determination. A flow rate of 0.5 mL/min was

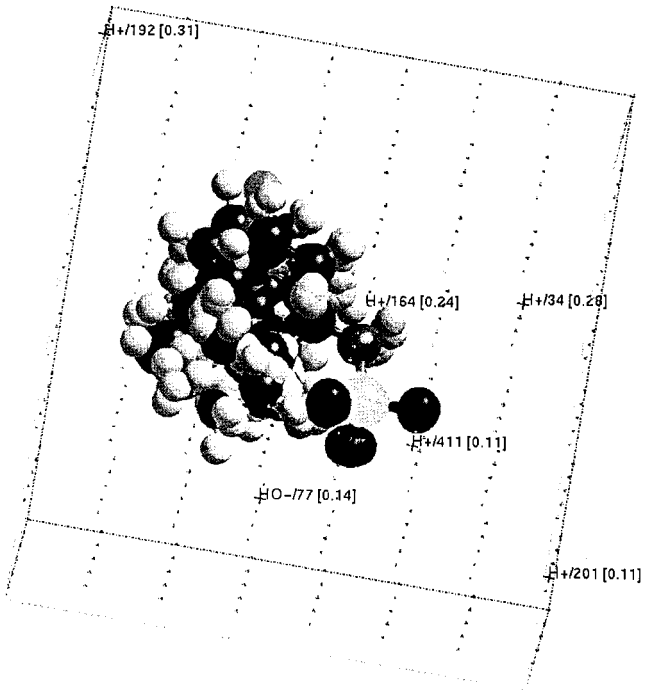

Figure 6. Results from the molecular field analysis. The molecules are overlapped via the $NH-SO_3$ moiety. The molecular field at all grid points was calculated. The field values at the six specific (illustrated) grid points were selected by the genetic algorithm as producing the best regression equation with the coefficients in square brackets.

maintained, and the eluent was water. Potassium bromide was used as a standard in each run, and k_w was calculated from the equation

$$k_w = (t_{RNHSO3Na} - t_{KBr})/t_{KBr}$$

where $t_{RNHSO3Na}$ and t_{KBr} are the retention times of the sulfamate under study and potassium bromide, respectively. **20** had an exceptionally long retention time, and use of the value of k_w (−1.005) led to a substantial deviation by this compound in the subsequent QSAR analysis. Thus, the value of k_w was calculated for **20** (see footnote, Table 2). A calculated value of k_w had to be used for **21** also because this compound was not available to us (see footnote, Table 2).

LITERATURE CITED

Acton, E. M.; Stone, M. Potential New Artificial Sweeteners from Study of Structure-Taste Relationships. *Science* **1976**, *193*, 584−586.

Alexander, E. R. Studies in the mechanism of the Sulfonation of Aromatic Amines. III Rearrangement of Barium Phenylsulfamates. *J. Am. Chem. Soc.* **1948**, *70*, 1274−1277.

Andrea, T. A.; Kalayeh, H. Applications of Neural Networks in Quantitative Structure−Activity Relationships of Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.

Aoyama, T.; Suzuki, Y.; Ichikawa, H. N. Neural Networks Applied to Quantitative Structure−Activity Relationship Analysis. *J. Med. Chem.* **1990**, *33*, 2583−2590.

Benson, G. A.; Spillane, W. J. Structure Activity Studies on Sulfamate Sweeteners. *J. Med. Chem.* **1976**, *19*, 869−872.

Braumann, T. Determination of Hydrophobic Parameters by reverse phase HPLC: Theory, Experimental Techniques and Applications in studies on Quantitative Structure Activity Relationships. *J. Chromatogr.* **1986**, *373*, 191−255.

CERIUS2 program, version 2.0; Molecular Simulations: Cambridge, U.K., 1996.

Daniel, J. R. In *Sweetness: Theory and Design Frontiers in Carbohydrate Research−1: Food Applications*; Millane, R. P., BeMiller, J. N., Chandorasekaran, R., Eds.; Elsevier: Amsterdam, 1989; pp 34−65.

Daniel, J. R.; Whistler, R. L. Sweetness-Structure Correlation in Carbohydrates. *Cereal Chem.* **1982**, *59*, 92−95.

Deutsch, E. W.; Hansch, C. Depdendence of Relative Sweetness on Hydrophobic Bonding. *Nature* **1966**, *211*, 75.

Friedman, J. *Multivariate Adaptive Regression Splines*; Technical Report 102; Laboratory for Computational Statistics, Department of Statistics, Stanford University: Stanford, CA, Nov 1988 (revised Aug 1990).

Greenberg, M. J. The Importance of Hydrophobic Properties of Organic Compounds and their taste Intenstities: A Quantitative Structure Taste Intensity Study. *J. Agric. Food Chem.* **1980**, *26*, 562−566.

Hansch, C.; Leo, A.; Taft, R. W., A Survey of Hammett Subsituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165−191.

Höltje, H.-D.; Kier, L. B. Sweet Taste Reception studies using Model Interaction Energy Calculations. *J. Pharm. Sci.* **1974**, *63*, 1722−1725.

Hurd, C. D.; Kharasch, N. The Dipole Ion Structure of the 2-thiazolylsulfamic Acids. *J. Am. Chem. Soc.* **1946**, *68*, 653−658.

Iwamura, H. Structure-Taste Relationship of Perillartine and Nitro-and Cyanoaniline Derivatives. *J. Med. Chem.* **1980**, *32*, 308−312.

Iwamura, H. Structure Sweetness Relationship of L-Aspartyl Dipeptide Analogues. A Receptor Site Topology. *J. Med. Chem.* **1981**, *24*, 572−583.

Kier, L. B. A Shape Index from Molecular Graphs. *Quantum Struct.-Act. Relat.* **1985**, *4*, 109−116.

Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Chemistry and Drug Research*; Medicinal Chemistry Monograph 14; De Stevens, G., Ed.; Academic Press: New York, 1976.

McFarland, J. W. In *On Understanding of Drug Potency. Progress in Drug Research*; Jucker, E., Ed.; Birkhäuser Verlag: Basel, 1971; Vol. 15, pp 123−126.

Miyashita, Y.; Takahasha, Y.; Takayama, C.; Sumi, K.; Nakatsuka, K.; Ohkubo, T.; Abe, H, Sasaki, S.-i. Structure-Taste Correlation of L-Aspartyl Dipeptides Using SIMCA methods. *J. Med. Chem.* **1986a**, *29*, 906−912.

Miyashita, Y.; Takahashi, Y.; Takayama, C.; Ohkubo, T.; Funatsu, K.; Sasaki, S.-i. Computer Assisted Structure/Taste Studies on Sulfamates by Pattern Recognition methods. *Anal. Chim. Acta* **1986b**, *84*, 143−149.

Miyashita, Y.; Kanaya, S.; Katsumi, H.; Takayama, C.; Nagakura, A.; Sasaki, S.-i. Structure-taste correlation of substituted β-(3-hydroxy-4-methoxyphenyl)ethylbenzenes using pattern recognition method. *Chem. Senses* **1989**, *14*, 781−792.

MOPAC, version 6. J. J. P. Stewart; obtained as QCPE 455, Quantum Chemistry Program Exchange, Indiana University.

Nofre, C.; Pautet, F. Sodium *N*-(cyclopentylmethyl) sulfamate; a new Synthetic Sweetener. *Naturwissenschaften* **1975a**, *97*, 62−69.

Nofre, C.; Pautet, F. Synthese de *N*-alkylsulfamates de sodium. *Bull. Chem. Soc. Fr.* **1975b**, 686−688.

Okuyama, T.; Miyashita, Y.; Kanaya, S.; Katsumi, H.; Sasaki, S.-i.; Randic, M. *J. Comput. Chem.* **1988**, *9*, 636−646.

Rogers, D. R.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure−activity relationships and quantitative structure−property relationships. *J. Chem. Inf. Comput. Sci* **1994**, *34*, 854−866.

So, S.-S.; Karplus, M. Evolutionary Optimization in Quantitative Structure−Activity Relationship: An Application of Genetic Neural Networks. *J. Med. Chem.* **1996**, *39*, 1521−1530.

Spillane, W. J.; McGlinchey, G. Structure−Activity Studies on Sulfamate Sweetness. II. Semiquantitative Structure-Taste relationships for Sulfamate ($RNHSO_3^-$) Sweeteners−the role of R. *J. Pharm. Sci.* **1981**, *70*, 933−935.

Spillane, W. J.; Sheahan, M. B. Semiquantitative and Quantitative Structure Taste Relationships for Carbo and Het-

erosulfamate ($RNHSO_3^-$) Sweeteners. *J. Chem. Soc., Perkin Trans. 2* **1989**, 741−746.

Spillane, W. J.; McGlinchey, G.; Muirceartaigh, I.; Benson, G. A. Structure Activity Studies on Sulfamates. III: Structure-Taste Relationships for Heterosulfamates. *J. Pharm. Sci.* **1983**, *72*, 852−856.

Spillane, W. J.; Sheahan, M. B.; Simmie, J. M.; Cunningham D.; McArdle, P.; Higgins, T. *J. Chem. Soc., Chem. Commun.* **1989**, 545−547.

Spillane, W. J.; Sheahan, M. B.; Ryder, C. A. Synthesis and taste properties of sodium disubstituted phenylsulfamates. Structure-taste relationships for sweet and bitter/sweet sulfamates. *Food Chem.* **1993**, *47*, 363−369.

Spillane, W. J.; Ryder, C. A.; Sheahan, M. Synthesis and taste properties of Sodium monosubstituted phenylsulfamates. *Food Chem.* **1994**, *57*, 405−411.

Spillane, W. J.; Ryder, C. A.; Walsh, M. R.; Curran, P. J.; Concagh, D. G.; Wall, S. N. Sulfamate Sweeteners. *Food Chem.* **1996**, *56*, 255−262.

Takahashi, Y.; Miyashita, Y.; Tanaka, Y.; Abe, H.; Sasaki, S.-i. A Consideration for Structure-Taste Correlations of Perillartines Using Pattern Recognition Techniques. *J. Med. Chem.* **1982**, *25*, 1245−1248.

Takahashi, Y.; Miyashita, Y.; Tanaka, Y.; Abe, H.; Sasaki, S.-i. Discriminative Structure Analysis using Pattern Recognition Techniques in the Structure-Taste Problem of Perillartines. *J. Pharm. Sci.* **1984**, *73*, 737−741.

Taylor, P. J.; Wait, A. R. σ' values for heterocycles. *J. Chem. Soc., Perkin Trans. 2* **1986**, 1765−1770.

Topliss, J. G.; Edwards, R. P. Chance factors in studies of quantitative structure−activity relationships. *J. Med. Chem.* **1979**, *22*, 1238−1244.

TSAR v 2.4. Tools for Structure Activity Relationships, Oxford Molecular Ltd., 1996.

Van der Heijden, A.; Brussel, L. B. P.; Peer, H. G. Quantitative Structure Activity Relationships (QSARs) in sweet aspartyl dipeptide methyl esters. *Chem. Senses Flavour* **1979**, *4*, 141−152.

Van der Heijden, A.; Van der Wel, H.; Peer, H. G. Structure Activity Relationships in Sweeteners. Nitroanilines, sulfamates, oximes, isocoumarins, and dipeptides. *Chem. Senses* **1985a**, *10*, 57−72.

Van der Heijden, A.; van der Wel, H.; Peer, H. G. Structure−activity relationships in sweetness. II. Saccharins, acesulfames, chlorosugars, tryptophans and ureas. *Chem. Senses* **1985b**, *10*, 73−85.

Vaugnat, A. M.; Wagner, E. C. Vibrational Spectra and Structure of Solid Sulfamic Acid and the Sulfamate Ion. *J. Chem. Phys.* **1957**, *26*, 77−82.

Zalewski, R. I. Structure-Sweet taste relationships by discriminant and principal component analysis. *J. Chem. Phys. Phys-Chem. Biol.* **1992**, *89*, 1507−16.

Received for review January 28, 1998. Revised manuscript received May 18, 1998. Accepted May 20, 1998. We thank the EC for support (AIR3 Project CT94-2107), and C.A.R. acknowledges a maintenance grant from the Irish Government.

JF980095C