5486 J. Agric. Food Chem. 2009, 57, 5486–5493
DOI:10.1021/jf9002472

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Development of Structure−Taste Relationships for Thiazolyl-, Benzothiazolyl-, and Thiadiazolylsulfamates

William J. Spillane,* Catherine M. Coyle, Brendan G. Feeney, and Emer F. Thompson

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

A total of 28 new five-membered aromatic ring thiazolyl-, benzothiazolyl-, and thiadiazolylsulfamates, as their sodium salts, have been synthesized and combined with 30 known similar heterocyclic sulfamates to create a database for the study of structure−activity (taste) relationships (SARs) in this heterocyclic subgroup, which is known to contain a somewhat disproportionate number of sweet compounds compared to other groups of tastants. A series of nine parameters (descriptors) to describe the properties of the sulfamate anions were calculated in Spartan Pro and HyperChem programs. These are the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), length of the molecule, dipole moment, area, volume, E_{solv}, σ (from the literature), and log P. The taste data for all 58 compounds were categorized into three classes, namely, sweet (S), nonsweet (N), and nonsweet/sweet (N/S). Discriminant analysis only classified 44 of the 58 compounds correctly. Classification and regression tree analysis (CART) using the S_ Plus program proved highly effective, in that the derived tree correctly classified 46 compounds from a training set of 48 and, from a computer randomly selected test set of 10 compounds, 7 had their taste correctly predicted. A second tree was grown using the additional taste category N/S, and this tree also performed extremely well, with 8 of the 10 compounds in the test set correctly classified. These trees should be very reliable for predicting the tastes of other heterocyclic sulfamates, which belong to the subset used here.

KEYWORDS: Heterosulfamates; tastants; CART analysis; sweetness; SARs

INTRODUCTION

Since 1946, about 10 groups have been active worldwide in synthesizing heterosulfamates, i.e., sulfamates that contain one or more heteroatoms in R in $RNHSO_3Na$, in a search for new sweeteners (*1*). By 1983, 33 new heterosulfamates were known (*2*); by 1989, 56 compounds had been made (*3*), and that number almost doubled in the next decade to 101 (*4*); and by 2003, 132 compounds had been synthesized and tested for taste (*1*). Further details with regard to all 132 compounds including the locations of the groups (*1*) and (if available) full taste portfolios can be accessed (*1*−*5*). The first sweet heterosulfamate compounds **4**−**6** and **8** (**Figure 1**) were made by Hurd and Kharasch in 1946 (*5*), and their sweetness was discovered by chance. Probably, the authors, who were studying the rearrangement of the sulfonate groups in the thiazolylsulfamates, were prompted to taste them because of a contemporaneous report of the sweetness of various non-heterosulfamates.

The numbering system used in **Figure 1** was introduced initially for the first 33 compounds in 1983 (*2*) and continued as new compounds were made by the various groups. For the 132 heterosulfamates, both open-chain and cyclic, representing many different classes of heterocompounds, some limited structure−taste relationships (SARs) were developed using a series of descriptors, the mathematical techniques of linear (LDA) and quadratic (QDA) discriminant analyses, and classification and regression tree (CART) analysis (*4*). Various descriptors that took the account of spatial, electronic, and other factors were employed. LDA and QDA performed poorly, correctly classifying only 70 and 68%, respectively, into sweet and nonsweet categories, and the CART analysis was somewhat more successful, classifying 81% of the compounds into the correct categories. Bearing in mind the large diversity of structural types in the database, these rather poor classification rates were not surprising.

It was noticed that certain classes of sulfamate tastants displayed little or no sweetness, while other classes seemed to be more likely to show some degree of sweet taste. For example, although 18 sulfamates containing the pyridine ring system have been synthesized, only 5 displayed any sweetness. In contrast, sulfamates containing the thiazole, benzothiazole, or thiadiazole ring systems tended to display sweetness more frequently. It was therefore decided to combine these three subsets and seek structure−taste relationships for the combined subsets. Such compounds all contain a five-membered core aromatic ring with one nitrogen and one sulfur atom (thiazoles and benzothiazoles) or two nitrogen atoms and one sulfur atom (thiadiazoles).

*To whom correspondence should be addressed. Telephone: +353-91-492475. Fax: +353-91-525700. E-mail: william.spillane@nuigalway.ie.

 Published on Web 05/20/2009

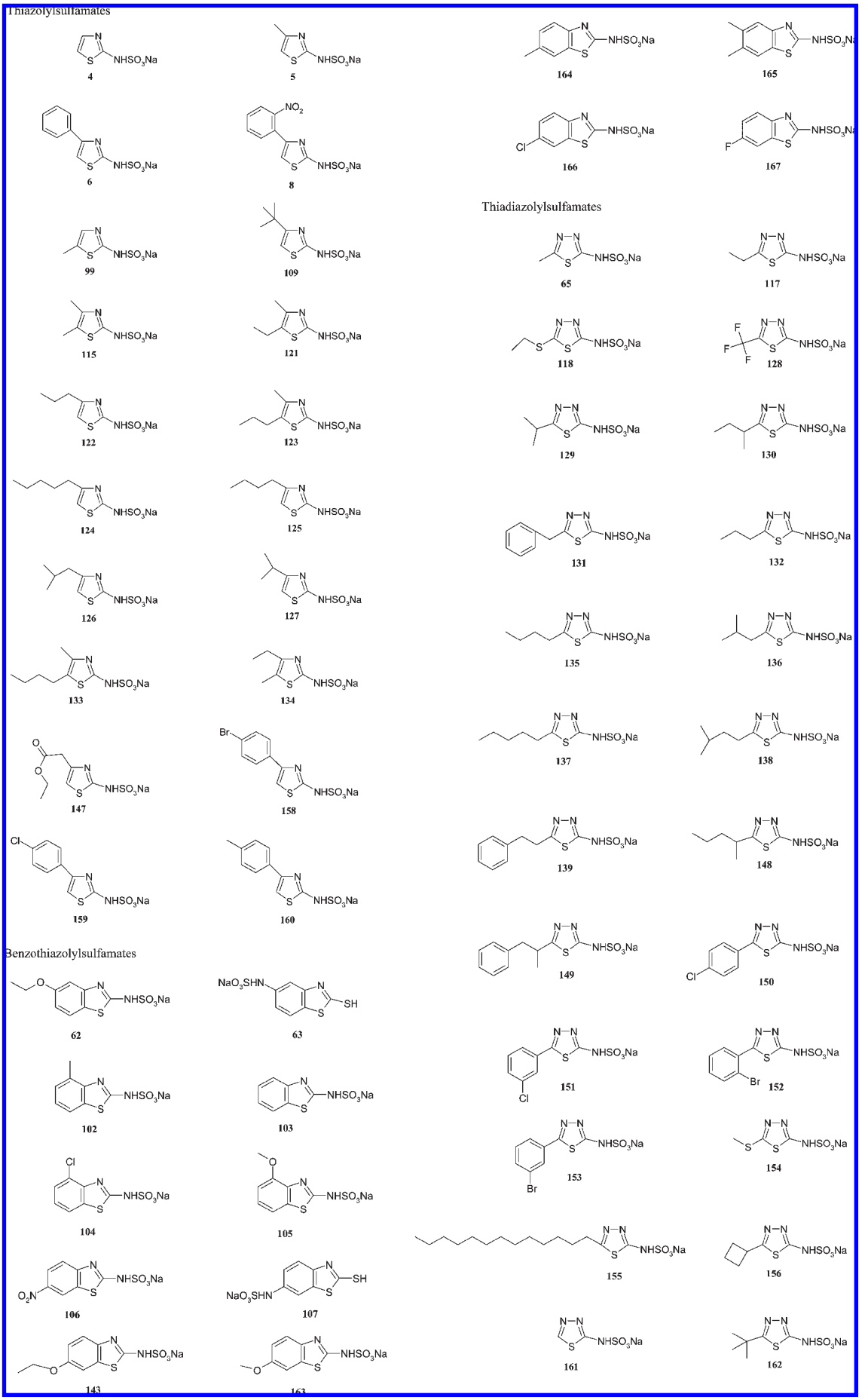

Figure 1. Structures and numbers of thiazolyl-, benzothiazolyl-, and thiadiazolylsulfamates.

Only 30 of the compounds in the enlarged set of 132 fell into this subset, and a program of synthesis over a few years almost doubled the number of compounds to 58. Thus, in this present work, 28 new sulfamates (and 3 new precursor amines) have been made. The objective of this study was to find a structure−taste relationship for the full set of 58 compounds.

MATERIALS AND METHODS

Chemistry. Most of the precursor amines required for the synthesis of the sulfamates in this work were commercially available. Three amines, which are new compounds, were synthesized by standard methods. These are the precursor amines for the thiadiazolylsulfamate numbers **148**, **149**, and **156** (**Figure 1**). The precursor amines for the thiazolylsulfamate compounds **133** and **134** are 4-methyl-5-*n*-butyl-2-aminothiazole (**133A**) and 4-ethyl-5-methyl-2-aminothiazole (**134A**), respectively, and they were synthesized by the reaction of thiourea and iodine with hexan-2-one and pentan-2-one, respectively, using the methods of Dodson and King (*6*) and King and Hlavacek (*7*). Both of these amines are known. When synthesis is carried out with unsymmetrical ketones, as in this case, two different iodoketones are formed *in situ* and these will in turn give rise to two isomeric thiazoles. Starting from hexan-2-one, thiazole **133A** forms and the other isomer would be 4-*n*-butyl-5-methyl-2-aminothiazole. However, this was not formed in sufficient quantities to sulfamate it. Starting from pentan-2-one, thiazole **134A** is formed together with thiazole **121A**, and upon sulfamation, this gives **121** (see **Figure 1**). The two isomers formed in each reaction were separated by flash chromatography on a silica gel column using 40−60 °C petroleum spirit/diethyl ether (50:50, v/v). Crude **133A** gave a pale yellow liquid that solidified upon standing to give 45% yield. ^{1}H nuclear magnetic resonance (NMR) (DMSO-d_6) δ: 0.85 (t, 3H, H10a,b,c), 1.26 (m, 2H, H9a,b), 1.40 (m, 2H, H8a,b), 1.94 (s 3H, H5a,b,c), 2.46 (t, 2H, H7a,b), 6.51 (s, 2H, NH_2). ^{13}C NMR δ: 13.64 (C10), 14.49 (C9), 21.41 (C8), 25.11 (C7), 33.47 (C5), 117.84 (C6), 141.66 (C4), 164.50 (C2). Crude **134A** was collected by suction filtration and recrystallized from *n*-hexane giving yellow crystals in 67% yield. ^{1}H NMR (DMSO-d_6) δ: 1.04 (t, 3H, H6a,b,c), 2.07 (s, 3H, H8a,b,c), 2.31 (q, 2H, H5a,b), 6.51 (s, 2H, NH_2). ^{13}C NMR δ: 10.20 (C6), 13.77 (C8), 21.24 (C5), 110.87 (C7), 147.77 (C4), 1764.41 (C2). The NMR data confirm that the correct thiazole isomers were used in the synthesis of **133** and **134**. For **134**, its isomer is **121** and examination of the NMR data for both precursor amines, i.e., **134A** and **121A**, clearly identifies each. For **133**, the alternative isomer was not available but the use of a NMR simulation program confirmed the structure of **133A** and showed that the alternative isomer would display a different shift pattern.

The precursor amines for thiadiazolylsulfamate numbers **148**, **149**, and **156** are 2-amino-5-(methylbutyl)-1,3,4-thiadiazole (**148A**), 2-amino-5-(1-methyl-2-phenylethyl)-1,3,4-thiadiazole (**149A**), and 2-amino-5-cyclobutyl-1,3,4-thiadiazole (**156A**). The synthesis of the three thiadiazoles was based on the procedure of Chubb and Nissenbaum (*8*), in which a carboxylic acid in sulfuric acid is reacted with thiosemicarbazide. Compound **148A** was obtained as a white powder, mp 148−149 °C in 16% yield. ^{1}H NMR (DMSO-d_6) δ: 0.846 (t, 3H, H9a,b,c), 1.192−1.209 (m, 5H, H10a,b,c, H8a,b), 1.494−1.502 (m, 2H, H7a,b), 2.987−3.005 (m, 1H, H6a), 6.986 (s, 2H, NH_2). ^{13}C NMR (DMSO-d_6) δ: 13.69 (C9), 19.67 (C8), 20.90 (C7), 34.87 (C10), 38.99 (C6), 163.18 (C5), 167.73 (C2). Compound **149A** formed as a beige powder, decomposing at 170 °C in 16% yield. ^{1}H NMR (DMSO-d_6) δ: 1.190 (d, 3H, H14a,b,c), 2.490 (s,1H, H6a), 2.801 (d,d, 1H, H7a), 2.939 (d,d 1H, H7a), 6.972 (s, 2H, NH_2), 7.01−7.236 (m, 5H, H9a, H10a, H11a, H12a, H13a). ^{13}C NMR (DMSO-d_6) δ: 20.44 (C14), 36.83 (C7), 42.47 (C6), 126.12 (C10, C11, C12), 128.15 (C9, C13), 128.96 (C8), 139.24 (C5), 162.93 (C2). Compound **156A** was synthesized as a beige powder, mp 226.5−227 °C in 85% yield. ^{1}H NMR (DMSO-d_6) δ: 1.217 (s, 1H, H8a), 1.844 (m, 1H, H8b), 1.942 (m, 1H, H6a), 2.130 (m, 2H, H7a,b), 2.296 (m, 2H, Ha,b), 7.043 (s, 2H, NH_2). ^{13}C NMR (DMSO-d_6) δ: 17.97 (C8), 28.93 (C7, C9), 34.74 (C6), 162.14 (C5), 167.83 (C2). Amines **148A** and **149A** gave C, H, and N microanalysis within at least ±0.5%, but amine **156A** gave the following: theoretical C, 46.45%; H, 5.87%; N, 27.10%; found C, 46.37%; H, 5.26%; N, 26.92%.

The 28 new sulfamates in this current work have been synthesized by literature methods (*9*, *10*), involving the addition of chlorosulfonic acid to pyridine at ∼0 °C and isolation in all cases of the sodium salt of the heterocyclic sulfamate. The detailed procedures involved in the synthesis, isolation, and purification of aromatic sulfamates, such as those discussed here, have been described in detail in recent reports (*11*, *12*). All of the new sodium salts gave a good clean sulfamate test, were free of sulfate, gave the characteristic IR frequencies corresponding to the $-NHSO_3^-$ group, and gave C, H, and N microanalysis within at least ±0.5% and generally within ±0.3%, except compounds **158**, **160**, **163**, **164**, and **166**. Their analyses are as follows: compound **158**, theoretical C, 30.26%; H, 1.69%; N, 7.84%; found C, 29.97%; H, 0.86%; N, 7.90%; compound **160**·1.75H_2O, theoretical C, 37.09%; H, 3.89%; N, 8.65%; found C, 37.58%; H, 3.14%; N, 8.04%; compound **163**·0.66H_2O, theoretical C, 32.65%; H, 2.85%; N, 9.52%; found C, 32.25%; H, 1.95%; N, 10.01%; compound **164**·0.75H_2O, theoretical C, 34.34%; H, 3.06%; N, 10.01%; found C, 34.32%; H, 2.36%; N, 8.50%; **166**·1H_2O, theoretical C, 27.59%; H, 1.98%; N, 9.19; found C, 27.62%; H, 1.86%; N, 8.06%. Their purity was confirmed by NMR analysis.

Instrumentation. ^{1}H and ^{13}C NMR spectra were recorded in DMSO-d_6 on a JEOL 400 MHz spectrometer. IR analysis was carried out using a Perkin-Elmer FTIR spectrum 1000, and C, H, and N elemental microanalysis was performed on a Perkin-Elmer 2400 series II analyzer. The pH meters used were Oakton pH 6 Acorn series and Jenway model 3310 buffered at pH 4 and 7.

Programs. Spartan Pro 04 (supplier Wavefunction, Inc.) and HyperChem Pro 6 (supplier ScienceSoftware) software were used in the calculation of eight of the nine descriptors. The Minitab 15 (supplier Minitab, Inc.) program was used to carry out LDA and QDA and to analyze the collinearity of the descriptors. CART analysis was carried out using the S_Plus 8.0 (supplier TIBCO Software, Inc.) program. Cross-validation was carried out within each program. For the simulated NMRs, the ACD Laboratories NMR simulation program (supplier Advanced Chemistry Development, Inc.) was used. All programs used in this work were available in the Computer Services Department at National University of Ireland (NUI), Galway.

Sensory Analysis of Heterocyclic Sulfamates. Details of the four primary standards (sucrose, citric acid, quinine sulfate, and sodium chloride) used have been given previously (*11*, *12*). The "sip and spit" method was used for taste assessment. All assessors were volunteers, were briefed on the nature of the work, and were extensively trained using the four standards, distilled water, and at times, a dilute sodium cyclamate solution to familiarize the assessors with the particular quality of sweetness found in sulfamates. Taste profiles on those heterocyclic sulfamates previously prepared prior to this work are available in the literature (*1*, *4*, *5*). Taste data for, first, 18 of the new compounds are given in **Table 1**, including a final predominant taste, i.e., N, nonsweet, S, sweet, and N/S, nonsweet/sweet, describing compounds that have clear nonsweetness (usually either bitterness, blandness, or tastelessness) and definite sweetness, albeit weak. The percents given in **Table 1** are arrived at as follows: for example, thiazolylsulfamate **134** was tasted by 7 assessors, with one of them (100% × 1/7) finding the compound to be bitter (14%), five (100% × 5/7) finding it to be sour (72%), and no other tastes detected. Thus, each taste detected by a panelist merits 100%/7, that is, 14%. This method of assessment can lead to a total percentage greater than 100%; for example, thiadiazolylsulfamate **136** was also tasted by seven assessors and found to be 28% sweet (100% × 2/7 assessors), 28% sour (100% × 2/7), 14% bitter (100% × 1/7), 28% salty (100% × 2/7), and 28% tasteless (100% × 2/7). The remaining 10 compounds were examined more recently by one of us (W.J.S.). This is necessary nowadays because of ethical guidelines.

Full taste profiles for 18 of the sulfamates were determined by panels of assessors some time ago, and the *modus operandi* employed has been described in detail elsewhere (*11*, *12*). The 10 compounds that were tasted by one taster only are the thiazolylsulfamates **158**−**160**, the benzothiazolylsulfamates **163**−**167**, and the thiadiazolylsulfamates **161** and **162**.

RESULTS AND DISCUSSION

In **Figure 1**, the structures of all 58 compounds belonging to the thiazole/benzothiazole/thiadiazole subsets are shown. Those made prior to this work bear numbers between **4** and **132**, and the 28 new compounds have numbers falling between **133** and **167**. Some compounds between **133** and **167** are missing from

Table 1. Taste Profiles[a] for 28 New Sulfamates

sulfamate number[b]	pH	% sweet	% sour	% bitter	% salty	% tasteless	predominant taste (N, S, or N/S)[c]
				Thiazole			
133	11.0	0	14	86	0	0	N
134	11.3	0	72	14	0	0	N
147	9.11	0	14	85	0	0	N
158	7.46		bitter, some sweetness like aniseed, silky texture				N/S
159	7.37		bitter, subdued perfume like taste				N
160	6.53		very bitter and then faint sweetness				N/S
				Benzothiazole			
143	11.5	0	0	100	0	0	N
163	6.83		bland				N
164	6.85		very faint sweetness				N
165	7.98		some sweetness, some bitterness				N/S
166	7.89		delayed sweetness that then becomes stronger				S
167	7.74		very bitter, lingers strongly				N
				Thiadiazole			
135	9.90	42	14	42	0	0	N
136	9.65	28	28	14	28	28	N
137	9.30	28	0	28	42	14	N
138	9.86	14	14	28	28	58	N
139	9.80	0	58	14	14	0	N
148	9.82	0	0	100	0	0	N
149	10.71	0	0	100	0	0	N
150	10.41	0	43	14	43	29	N
151	9.86	29	0	71	14	14	N
152	10.4	86	0	85	29	0	N/S
153	11.8	86	0	71	0	0	N/S
154	9.71	0	0	100	0	28	N
155	11.1	100	14	14	14	0	S
156	10.8	57	0	28	14	0	S
161	6.88		pleasant slight sweetness				S
162	8.25		delayed sweetness				S

[a] All sodium sulfamates were tasted as 0.01 M solutions in freshly distilled water at temperatures of 20.5 ± 1 °C. Solutions were tasted within 24 h of preparation. Between 6 and 8 assessors were used, except for those sulfamates where a written description of the taste(s) is recorded; in these cases, one assessor was used. [b] The sulfamate number refers to the structures presented in **Figure 1**. [c] N, nonsweet; S, sweet; and N/S, both sweetness and nonsweetness detected.

Figure 1. This is because they are heterocompounds, which do not fall into the combined subsets of current interest.

Descriptors Used. Nine parameters were employed as descriptors for each of the tastants in this work to try to find a reliable qualitative structure−activity relationship (QSAR) for the group of heterocyclic compounds under study. They are shown in the heading to **Table 2**, and the values for each of the 58 compounds in the study are given in **Table 2**. Seven of descriptors were calculated using the Spartan Pro 04 software program. These descriptors are the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), length of the molecule, dipole moment, area of the molecule, volume of the molecule, and E_{solv}. Models of the compounds were first built within the program, and the lowest energy of conformation was found. A charge of −1 was assigned to the anionic sulfamates, ignoring the positive sodium cation, because this was the form of the molecule present in aqueous solution. The Hammett σ values were available from a standard compilation (see footnote e in **Table 2**), and where two substituents were involved, the algebraic sum of the two σ values was taken. Differing σ values are available for the *ortho*, *meta*, and *para* positions in six-membered aromatic rings, and there are numerous examples of these being employed to correlate electronic effects in other aromatic rings, including five-membered rings. These were, thus, systematically assigned to the substituents on the carbon atoms of the thiazolyl-, benzothiazolyl-, and thiadiazolylsulfamates according to the scheme shown in **Figure 2**. Thus, in the thiazoles, the 4 and 5 positions were assigned as *para* and *meta*, respectively; in the benzothiazoles, the 5, 6, and 7 positions were assigned as *ortho*, *meta*, and *para*, respectively; and in the thiadiazoles, only the 5 position can carry a substituent and this position was assigned as a *meta* one. For four of the compounds, no σ value was available; therefore, an estimation of its value was made. The σ value for the closest possible substituent to the substituent in the compound was used. For 4-ethylacetoxythiazolylsulfamate (**147**), the σ value for 4-acetoxy was used (−0.45). For 5-*n*-tridecylthiadiazolylsulfamate (**155**), the σ value of 5-*n*-heptyl was used (−0.07). The σ value of 5-*n*-propylphenyl was used instead of 5-*i*-propylphenyl for compound **149** (−0.12), and for 5-(2-bromo)phenylthiadiazolylsulfamate (**152**), the σ value for the 5-(3-bromo)phenyl substituent was used (0.09). Partition coefficients, log *P*, were calculated using the HyperChem Pro 6 software program.

HOMO and LUMO are electronic parameters that show electron-rich regions of the molecule, such as, lone pairs and regions susceptible to nucleophilic attack, respectively. Length, area, and volume are various measures of size and bulkiness and are known to be very important in relation to the R moiety in $RNHSO_3Na$ (*1*, *13*). The dipole moment gives a measure of bond polarity and charge separation throughout the molecule. E_{solv} is a measure of the energies of aqueous solvation and formation. This term provides information on hydrophobic/hydrophilic interactions for each compound. The Hammett σ values are

Table 2. Taste Data and Parameters for Heterocyclic Sulfamates

sulfamate number[a]	taste[b]	HOMO (eV)[c]	LUMO (eV)[c]	length (Å)[c,d]	dipole moment (debyes)[c]	area (Å^2)[c]	volume Spartan (Å^3)[c]	E_{solv} (kcal mol^{-1})[c]	σ[e]	log P[f]
					Thiazolylsulfamates					
4[g]	S	−5.25	3.38	4.962	10.24	159.92	128.69	−199.535	0.00	1.12
5[g]	S	−5.21	3.36	6.504	14.32	180.42	147.05	−206.453	−0.17	1.14
6[g]	N	−5.33	2.28	9.230	20.60	241.05	212.41	−172.242	−0.01	2.09
8[g]	N	−5.66	0.96	10.437	19.66	266.98	234.11	−169.834	0.20	−2.59
99[h]	N	−5.20	3.28	6.256	13.67	180.35	146.96	−207.254	−0.07	1.42
109[i]	N	−5.22	3.34	7.323	20.53	234.45	200.86	−215.797	−0.20	2.80
115[i]	S	−5.41	3.26	6.507	16.44	199.43	165.04	−213.835	−0.24	1.45
121[i]	N	−5.16	3.26	6.949	18.23	218.88	183.47	−218.873	−0.24	1.92
122[i]	S	−5.23	3.34	8.522	20.04	220.67	183.77	−217.841	−0.13	2.28
123[i]	N	−5.18	3.25	7.116	19.38	236.87	201.59	−224.478	−0.23	2.31
124[i]	S	−5.24	3.33	10.518	25.35	259.95	220.43	−230.394	−0.15	3.08
125[i]	S	−5.24	3.33	9.283	22.63	240.80	202.11	−224.500	−0.16	2.68
126[i]	S	−5.24	3.33	8.491	21.30	238.58	201.71	−221.570	−0.12	2.61
127[i]	N	−5.22	3.34	7.341	19.41	217.92	183.29	−215.083	−0.15	2.29
133[j]	N	−5.19	3.24	8.055	20.99	256.92	219.93	−231.060	−0.25	2.71
134[j]	N	−5.17	3.26	7.038	18.18	219.24	183.41	−218.795	−0.21	1.92
147[j]	N	−5.62	2.88	7.435	14.69	247.03	213.13	−290.474	-0.45	1.17
158[j]	N/S	−5.47	1.85	11.082	18.65	260.92	230.34	−167.490	0.12	2.14
159[j]	N	−5.44	1.94	10.905	18.91	256.16	225.74	−178.923	0.12	1.87
160[j]	N/S	−5.31	2.23	10.699	24.14	260.92	230.61	−179.744	−0.03	2.24
					Benzothiazolylsulfamates					
62[h]	N	−5.49	2.62	10.742	22.90	258.34	226.15	−227.956	0.10	1.00
63[h]	N	−5.35	1.97	9.136	14.62	222.96	196.25	−183.789	0.25	0.87
102[i]	S	−5.51	2.68	7.468	17.44	227.37	198.41	−190.859	−0.13	1.80
103[i]	N	−5.54	2.68	7.467	15.75	207.81	180.24	−184.572	0.00	1.65
104[i]	N	−5.67	2.44	7.478	15.11	222.91	193.53	−189.802	0.68	1.43
105[i]	S	−5.57	2.59	7.486	16.23	237.43	207.77	−219.868	0.00	0.66
106[i]	N	−6.16	1.29	8.717	11.27	233.83	201.95	−179.747	0.78	−3.03
107[i]	N	−5.29	2.14	9.227	13.22	223.95	196.36	−187.016	0.15	0.87
143[j]	N	−5.49	2.59	10.753	23.81	258.76	226.30	−227.373	−0.24	1.00
163[j]	N	−5.46	2.64	9.737	20.98	237.66	207.60	−222.717	−0.27	0.66
164[j]	S	−5.48	2.63	8.725	19.03	227.62	198.42	−191.919	−0.17	1.80
165[j]	N/S	−5.47	2.63	8.883	21.72	245.35	216.08	−198.666	−0.24	1.96
166[j]	S	−5.68	2.36	8.915	14.07	222.99	193.60	−190.188	0.23	1.43
167[j]	N	−5.64	2.39	8.600	14.07	213.88	185.16	−227.361	0.06	1.05
					Thiadiazolylsulfamates					
65[h]	S	−5.59	2.97	6.284	11.19	174.74	140.08	−186.896	−0.07	4.67
117[i]	S	−5.59	2.98	7.185	14.54	194.90	158.43	−192.680	−0.07	5.30
118[i]	S	−5.51	2.50	7.406	14.74	215.76	176.81	−191.566	0.18	3.90
128[i]	S	−6.14	2.26	7.224	6.95	192.74	154.69	−329.595	0.43	3.70
129[i]	S	−5.59	2.99	7.010	16.14	212.24	176.28	−205.800	−0.04	5.86
130[i]	N	−5.62	2.97	7.233	18.04	233.10	194.78	−201.952	−0.08	6.25
131[i]	N	−5.67	2.44	9.147	20.18	256.88	224.08	−160.027	−0.08	4.44
132[i]	S	−5.60	2.97	7.232	16.52	214.32	176.74	−198.263	−0.06	5.69
135[j]	N	−5.74	2.83	8.808	22.08	233.77	195.28	−205.673	−0.08	6.09
136[j]	N	−5.61	2.95	7.200	17.52	229.27	194.23	−201.197	−0.07	3.95
137[j]	N	−5.61	2.96	8.452	20.14	253.34	213.33	−211.156	−0.08	3.75
138[j]	N	−5.61	2.96	8.876	20.52	248.63	212.69	−206.395	-0.08	5.86
139[j]	N	−5.64	2.46	8.876	20.51	276.18	242.42	−166.542	−0.07	4.83
148[j]	N	−5.61	2.98	7.493	18.83	251.79	213.01	−207.871	-0.10	4.34
149[j]	N	−5.72	2.42	9.176	20.71	291.09	260.05	−168.719	-0.12	4.63
150[j]	N	−5.74	1.67	11.186	19.13	250.60	218.95	−161.434	0.15	3.30
151[j]	N	−5.68	1.84	10.157	16.01	250.39	218.78	−159.561	0.15	3.52
152[j]	N/S	−5.69	1.97	8.850	17.68	254.77	223.35	−146.146	0.09	3.79
153[j]	N/S	−5.69	1.81	10.314	15.88	255.17	223.38	−148.139	0.09	3.79
154[j]	N	−5.56	2.51	7.310	11.83	195.35	158.22	−185.501	0.15	3.56
155[j]	S	−5.62	2.95	16.398	43.29	393.15	341.70	−257.002	-0.07	7.58
156[j]	S	−5.59	2.96	7.639	17.48	219.26	185.78	−174.526	−0.05	4.85
161[j]	S	−5.68	3.05	4.841	7.78	152.07	121.64	−179.268	0.00	4.46
162[j]	N/S	−5.62	2.99	7.491	17.95	229.86	194.23	−197.408	−0.10	4.45

[a] The sulfamate number refers to a numbering system devised in early SARs to distinguish heterosulfamates in a simplified manner. [b] N, nonsweet; S, sweet; N/S, nonsweet/sweet. [c] Parameters were calculated using Spartan Pro 04 program. [d] The length was measured from the sulfur of the sulfamate unit to the furthest atom in the substituent. [e] The Hammett σ value is taken from Hansch, Leo, and Hoekman *Exploring QSAR—Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, D.C., 1995, and the σ value selected in each case was the most commonly used one, indicated by an asterisk next to the substituent in the book. [f] log P was calculated using HyperChem Pro 6 program. [g] Hurd and Kharasch (*5*). [h] Spillane et al. (*4*). [i] Spillane et al. (*1*). [j] Present work.

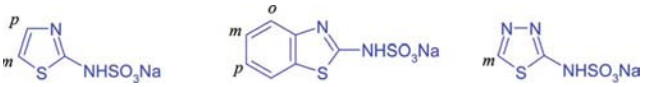

Figure 2. *ortho*, *meta*, and *para* positions assigned to thiazolyl-, benzothiazolyl-, and thiadiazolylsulfamates.

measures of inductive and resonance effects exerted by the differing substituents attached to the aromatic thiazole/benzothiazole/thiadiazole rings. Their usage here is most appropriate because they were originally devised as a means of assessing these effects in many aromatic systems. The partition coefficient P or log P is a measure of the hydrophilicity of a chemical substance. A high log P indicates that the compound is hydrophobic, while a low log P indicates a hydrophilic nature. Hydrophobicity and hydrophilicity are known to be important in devising SARs for many biological effects, and the special Hansch π parameter has been used widely for this. For the present sodium sulfamates, all of which are highly water-soluble and virtually insoluble in *n*-octanol, it is not possible to measure the partition coefficients experimentally and hence obtain the appropriate Hansch π values. Calculated P or log P values however do reflect the hydrophobic/hydrophilic interactions well.

The use of LDA and QDA on a large and more diverse group of 132 heterosulfamates had not proven very successful some years ago. LDA correctly classified only 70% of the compounds in the group, and QDA correctly classified about 68% (*1*). Notwithstanding these results, it was decided as a starting point to seek a structure−taste relationship using discriminant analysis again for the present group of heterocyclic compounds. This decision was motivated by the facts that the current data set of 58 sulfamates is much smaller than that looked at previously; it contains aromatic five-membered heterocycles *only* (the earlier set contained open-chain heterocompounds and both aromatic and non-aromatic heterocycles), and some of the descriptors used previously were subjective, in that they had to be physically measured, e.g., by assembling by hand space-filling Corey−Pauling−Koltun (CPK) models to obtain dimensions for the molecules. In the present work, eight of the nine descriptors used are computer-generated in the Spartan Pro and HyperChem programs and the remaining descriptor used Hammett σ values for the various substituents attached to the heterorings, can be obtained from the literature (see footnote e in **Table 2**).

Collinearity. When all of the data for the nine descriptors for each of the 58 compounds were inputted into the Minitab 15 program, a Pearson correlation matrix, shown in **Table 3**, was constructed by linear regression. This correlation matrix gives a measure of the relationship between each descriptor. If the degree of collinearity between two descriptors is too close, they may not be used together in predicting the classification functions for the discriminant analysis. A Pearson coefficient that tends toward unity is said to be too highly correlated, and if used together, these descriptors would yield inaccurate results in a statistical study and would not have a reliable predictive power. Ideally, therefore, the Pearson coefficient for any two descriptors should tend toward zero. In this study, only descriptors with Pearson coefficients less than ±0.300 were used together.

Cross-validation. Cross-validation is a method for testing the validity of results obtained in discriminant analysis. The program sets up a classification function for all of the observations from the supplied data. One observation, i.e., the data for one of the 58 compounds, is then removed, used as a test set, and is categorized by the classification function. This process is repeated until each observation has been removed and categorized. These results are all combined together to form error rates and are then applied to the original classification function. This type of cross-validation shows the accuracy of the classification function that has been obtained and allows for the best results to be achieved. It should be noted that the cross-validation procedure, which leaves out one compound at a time successively, *always* gives a lower percent fit, and the fact that this is only a few percentages lower in the discriminant analysis (DA) below is a good sign, indicating that the model is sound.

Table 3. Pearson Correlation Matrix of Heterocyclic Sulfamates for LDA and QDA

	HOMO	LUMO	length	dipole	area	volume	E_{solv}	σ
LUMO	0.562							
length	−0.125	−0.463						
dipole	0.243	0.129	0.747					
area	−0.110	−0.279	0.860	0.846				
volume	−0.115	−0.343	0.868	0.817	0.993			
E_{solv}	−0.080	−0.459	0.071	−0.107	−0.001	0.054		
σ	−0.570	−0.660	0.116	−0.367	−0.073	−0.048	0.256	
log P	−0.158	0.347	0.032	0.258	0.174	0.110	0.012	−0.258

LDA and QDA. LDA was carried out using all 58 sulfamates, i.e., S, N, and N/S. DA was used to discriminate between two properties, i.e., sweet and nonsweet; in this case, the six N/S compounds were divided up between these two categories. Five of the six N/S compounds were grouped with the S compounds because they had some element of sweetness, albeit with substantial amounts of another taste(s), and the remaining N/S compound number **160** was placed with the N group because it was "very bitter and then faint sweetness" (**Table 1**), portraying much more bitterness than sweetness. However, all other N/S compounds were deemed to have a stronger S than N quality. A series of classification functions were mathematically derived using between two and four descriptors for each subset. The "best" subset used the descriptors "area and HOMO"; however, it only predicted the tastes of 38 of the 58 compounds correctly (66%), and after cross-validation was carried out, the percent correctly predicted fell to 60%. Using three or four descriptors did not improve the categorization rate. Thus, the subset HOMO, area, and log P classified 36 compounds correctly (62%, which fell to 59% after cross-validation), and the subset with volume, Hammett σ, E_{solv}, and log P grouped 35 compounds (60% correctly, which dropped to 57% after cross-validation). QDA on all 58 compounds gave slightly better results. The subset that correctly classified the largest number of compounds contained LUMO, dipole moment, and log P, and it placed 44 compounds (76%) correctly. This dropped to 35 compounds (60%) after cross-validation.

Unfortunately, the derived classification function was skewed toward the N compounds, and although it correctly classified 91% of these, it only classified 56% of the S compounds. In summary, for these present tastants, both LDA and QDA performed poorly and the predictive ability of the best classification functions would be unreliable.

CART. CART analysis has become very popular over the last 20 years or so, and the technique has found application in many areas, including medicine, agriculture, food, and allied fields, and there are almost 300 applications in these areas. The successful classification of a diverse array of heterosulfamates (*1*) and simple aromatic mono- and disubstituted sulfamate tastants (*11*, *12*) has been achieved through the use of this technique. The high levels of classification obtained and reliable predictability have helped to popularise the method (see references in ref (*11*)) and prompted us to employ this method of analysis again in this current work. In this analysis, the descriptors are considered independently at each step and thus, unlike DA, collinearity is irrelevant. The technique of CART

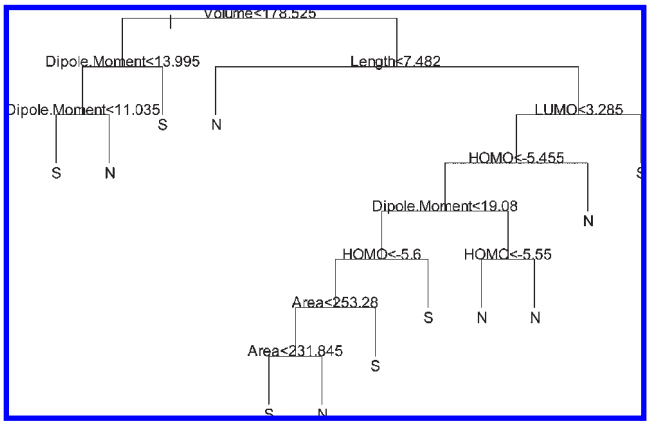

Figure 3. Classification tree for heterocyclic sulfamates: sweet (S) and nonsweet (N).

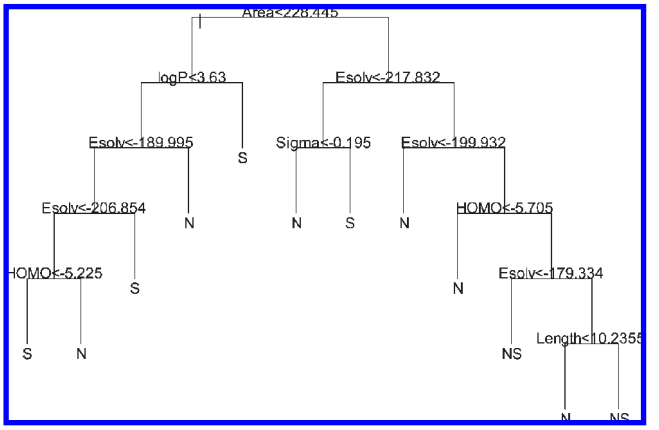

Figure 4. Classification tree for heterocyclic sulfamates: sweet (S), nonsweet (N), and nonsweet/sweet (N/S).

analysis was applied to a training set of 48 compounds, after the computer randomly selected 10 compounds as a test set. For the purpose of the analysis, the same five N/S compounds were again grouped with the sweet (S) compounds and the remaining N/S compound, **160**, was placed in the nonsweet (N) in the training set. The test set contained six nonsweet compounds and four sweet ones. They are numbers **8**, **62**, **127**, **129**, **135**, and **163** (all N) and **65**, **102**, **117**, and **155** (all S) in **Figure 1**. The technique grew a tree, shown in **Figure 3**, which correctly classified an impressive 46 compounds (95.8%). The use of a tree to carry out predictions has been illustrated previously with several examples (*11*, *12*). An example that illustrates its operation in the current work would be compound **133**, which is nonsweet (N), as determined by a taste panel (see the first entry in **Table 1**), and is also predicted to be nonsweet (N) using the tree in **Figure 3**. This prediction arises as follows: its volume is 219.93, which is greater than 178.525, and one moves to the right branch of the tree. The next node is length = 7.482. Compound **133** has a length greater than this (8.055), and thus, one moves to the branch on the right again. The next node is LUMO = 3.285. The LUMO value of **133** is less than this (LUMO = 3.24), and thus, one moves to the branch on the left. The next node is HOMO = −5.455. Compound **133** has a HOMO value of −5.19, which is greater than the node value on the tree, and thus, one moves to the right branch. This results in a terminal node of nonsweetness (N). This procedure is followed for each compound when reading the tree; i.e., if a value for the compound is less than the node value, one moves to the left, and if a Compound's value is greater than the node value, one moves to the right. The two misclassified compounds were both benzothiazolylsulfamates, namely, **165** (N/S but taken as S for this tree) and found to be nonsweet and **166** (S) but found to be nonsweet. Hence, all thiazolyl- and thiadiazolylsulfamates in the training set were correctly classified. In the test set, 7 of the 10 compounds were correctly classified. One benzothiazolylsulfamate, **102**, and two thiadiazolylsulfamates, **65** and **155**, were misclassified. It thus appears that the tree in **Figure 3** is, in general, reliable. The actual percentages for correct classification are 100% for thiazolyl-, 91.7% for thiadiazolyl-, and 78.6% for benzothiazolylsulfamates. For sweet compounds, the overall rate is 84% correctly classified, and for the nonsweet compounds, it is 97%.

A second classification tree was derived this time using the three categories N, S, and N/S and is shown in **Figure 4**. A total of 48 compounds were used in a training set to develop the tree, and 10 compounds (6 N, 3 S, and 1 N/S) randomly chosen were used in a test set. Although selected at random, this test set was not chosen by the computer. The computer-generated test set did not reflect the pool of compounds from the taste and structural points of view accurately, and thus, a more realistic test set was hand-picked. In the database, the ratio of S:N:N/S was 20:32:6 (total 58), but the computer selected a ratio of 5:1:4 (total 10) tastewise. The test set used consisted of compounds **8**, **121**, **130**, **138**, **139**, and **143** (all N), **102**, **126**, and **132** (all S), and **152** (N/S), which gives a ratio of 3:6:1 more in keeping with the actual distribution of the three tastes. The computer-selected ratio did not reflect the structural ratio either, which is 24:20:14 thiadiazoles, thiazoles, and benzothiazoles, respectively. A total of 3 of the 48 compounds from the training set were misclassified. Compounds **62** and **167** were both nonsweet and classified as sweet, while compound **158** was nonsweet and predicted to be N/S. This gives a classification rate of 93.75% for the tree. A total of 8 of the 10 compounds in the test set were classified correctly. Compound **8** was nonsweet and classified as N/S, while compound **152** was N/S and misclassified as N. This means that the test set was correctly classified at a rate of 80%. The overall classification rates for S, N, and N/S compounds were 100, 87.5, and 83.33%, respectively.

Both trees perform well in categorizing either N and S or N, S, and N/S combined, and either could be used to predict taste. The second tree (**Figure 4**) is slightly superior because it has 100% success in predicting sweetness, 87.5% for nonsweetness, and 83% for nonsweetness/sweetness and places 8 of the 10 test set compounds correctly. In contrast, the first tree (**Figure 3**) has 84% success for sweetness prediction and 97% for nonsweetness and correctly predicts the tastes of 7 of the 10 test set compounds. Which tree to use will be decided by the type of data available, i.e., N and S or N, S, and N/S, and either can be expected to offer accurate predictions of taste.

Two final points are worth noting. First, in the method employed here and in two other recent similar studies (*11*, *12*), a knowledge of the relative sweetness or other measures of sweetness are not required to apply the analysis. Second, the very successful use of more or less of the same nine descriptors here in this work and in the other studies to describe the properties of the compounds points to the possibility that these parameters are very important in seeking tastant and possibly other SARs. A limitation would be that, although it should be possible to predict within the heterocyclic subclass studied here whether or not a compound will be sweet (S), nonsweet (N), or nonsweet/sweet (N/S), one cannot predict the intensity of that taste.

ACKNOWLEDGMENT

Dr. Damien P. Kelly is thanked for advice on the use of the various programs involved.

LITERATURE CITED

(1) Spillane, W. J.; Kelly, L. M.; Feeney, B. G.; Drew, M. G. B.; Hattotuwagama, C. K. Synthesis of heterosulfamates. Search for structure−taste relationships. *ARKIVOC*, 2003, *7*, 297−309; available at http://www.arkat-usa.org/ark/journal/2003/107 McKervey/AM-736D/736D.pdf.

(2) Spillane, W. J.; McGlinchey, G.; O'Muircheartaigh, I.; Benson, G. A. Structure−activity studies on sulfamate sweeteners III: Structure−taste relationships for heterosulfamates. *J. Pharm. Sci.* **1983**, *72*, 852–856.

(3) Spillane, W. J.; Sheahan, M. B. Semi-quantitative and quantitative structure−taste relationships for carbo- and hetero-sulphamate ($RNHSO_3$) sweeteners. *J. Chem. Soc., Perkin Trans. 2* **1989**, 741–746.

(4) Spillane, W. J.; Ryder, C. A.; Curran, P. J.; Wall, S. N.; Kelly, L. M.; Feeney, B. G.; Newell, J. Development of structure−taste relationships for sweet and non-sweet heterosulfamates. *J. Chem. Soc., Perkin Trans. 2* **2000**, 1369–1374.

(5) Hurd, C. D.; Kharasch, N. The dipolar ion structures of 2-thiazolylsulfamic acids. *J. Am. Chem. Soc.* **1946**, *68*, 653–658.

(6) Dodson, R. M.; Carroll King, L. The reaction of ketones with halogens and thiourea. *J. Am. Chem. Soc.* **1945**, *67*, 2242–2243.

(7) Carroll King, L.; Hlavacek, R. J. The reaction of ketones with iodine and thiourea. *J. Am. Chem. Soc.* **1950**, *72*, 3722–3725.

(8) Chubb, F. L.; Nissenbaum, J. Some hypoglycemic thiadiazoles. *Can. J. Chem.* **1959**, *37*, 1121–1123.

(9) Boyland, E.; Manson, D.; Orr, S. F. D. The biochemistry of aromatic amines II. The conversion of arylamines into arylsulphamic acids and aryl-*N*-glucosiduronic acids. *Biochem. J.* **1957**, *65*, 417–423.

(10) Audreith, L. F.; Sveda, M. Preparation and properties of some N-substituted sulfamic acids. *J. Org. Chem.* **1944**, *9*, 89–101.

(11) Kelly, D. P.; Spillane, W. J.; Newell, J. Development of structure−taste relationships for monosubstituted phenylsulfamate sweeteners using classification and regression tree analysis (CART). *J. Agric. Food Chem.* **2005**, *53*, 6750–6758.

(12) Spillane, W. J.; Kelly, D. P.; Curran, P. J.; Feeney, B. G. Structure−taste relationships for disubstituted phenylsulfamate tastants using classification and regression tree (CART) analysis. *J. Agric. Food Chem.* **2006**, *54*, 5996–6004.

(13) Spillane, W. J.; McGlinchey, G. Structure−activity studies on sulfamate sweeteners II. Semi-quantitative structure−taste relationships for sulfamate ($RNHSO_3^-$) sweeteners—The role of R. *J. Pharm. Sci.* **1981**, *70*, 933–935.

Received January 21, 2009. Revised manuscript received April 21, 2009. The British Oxygen Co. is thanked for financial support.