JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Structure−Taste Relationships for Disubstituted Phenylsulfamate Tastants Using Classification and Regression Tree (CART) Analysis

WILLIAM J. SPILLANE,* DAMIEN P. KELLY, PATRICK J. CURRAN,† AND BRENDAN G. FEENEY

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

Forty-two new disubstituted phenylsulfamates have been synthesized, and 30 of these have been combined with 40 already available from earlier work to create a training database of 70 compounds. On the basis of panel taste data these were divided into three categories, N (nonsweet), N/S (nonsweet/sweet), and (S) sweet, and a "sweetness value" or weighting was also calculated for each compound. Using these 70 compounds as a training set and a series of nine predictors derived from Corey−Pauling−Koltun (CPK) models, calculated from the PC SPARTAN PRO program and Hammett σ values taken from the literature, a classification and regression tree analysis (CART) was carried out leading to a regression tree that correctly classified 62 of the 70 compounds (89% overall correct classification). The tree's predictive ability varies for the different taste categories, and for nonsweet compounds it is virtually 100%; for nonsweet/sweet compounds it is 66%, and for sweet compounds it is ≈75%. This tree correctly predicted taste categories for 10 compounds from a test set of 12 randomly selected from among the 42 new compounds (83% correct classification). Therefore, it can be used with a good degree of confidence to predict the tastes of disubstituted phenylsulfamates. For the design of new sweeteners, appropriate values or ranges of the descriptors are derived.

KEYWORDS: Tastants; cyclamates; CART analysis; phenylsulfamates; sweetness

INTRODUCTION

A first-generation sweetener, cyclamate [the generic name used to describe either individually or collectively sodium *N*-cyclohexylsulfamate (**Figure 1**), calcium *N*-cyclohexylsulfamate, and *N*-cyclohexylsulfamic acid], is still widely used almost worldwide and is an active competitor with other more recently developed alternative sweeteners. It has a number of important advantages (*1*) over various other sweeteners that have been launched, not the least being its good quality of sweetness both on its own and in synergy with other sweeteners. Another advantage arises from its favorable placing on the Birch scale. The latter has developed and explained an index of taste modality based on experimentally determined apparent specific volume (ASV) measurements ($cm^3\ g^{-1}$) (*2*). On the Birch scale salty molecules have ASVs in the range of ≈0.13−0.49, sweet molecules fall in the range of 0.55−0.68, sweet/bitter molecules fall in the range of 0.53−0.88, and a much wider range from 0.16 to 0.85 is found for sour molecules. For those sweet sulfamates for which ASVs have been measured (*3*) the values obtained lie within the "sweet" range given above; for example, sodium cyclamate has a value of 0.61, and two of the compounds studied here, namely, sodium 3,5-dimethyl- and 3,4-dimethoxyphenylsulfamates, have ASVs of 0.63 and 0.59, respectively.

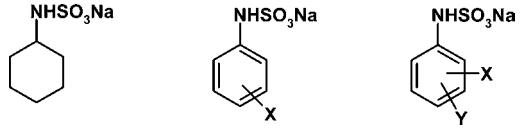

Figure 1. Molecular structures for (from left to right) sodium *N*-cyclohexylsulfamate (cyclamate), sodium *N*-*X*-phenylsulfamate, and sodium *X*,*Y*-phenylsulfamate.

Recently we have been looking at the effects on taste of changes in the sulfamate −$NHSO_3Na$ (*4*) moiety, and for many years several laboratories have been studying the structure−taste relationships of sulfamates $RNHSO_3Na$, where the R moiety has included alicyclic, aliphatic, heterocyclic, and hetero and aryl groups (*5*−*7*). In the case of the latter type we have recently published reliable structure−taste relationships for monosubstituted X-phenylsulfamates (*8*) (**Figure 1)** using over 80 compounds and a wide range of substituents (X) in the ortho-, meta-, and para-positions.

Some years ago we attempted with little success to find a reliable structure−taste relationship for a rather small set of 40 disubstituted X,Y-phenylsulfamates (**Figure 1**) (*9*) using Corey−Pauling−Koltun (CPK) parameters, Hammett σ values, and first-order molecular connectivity, the latter parameter being

* Corresponding author (e-mail william.spillane@nuigalway.ie).
† Present address: Schering-Plough Research Institute, Cambridge, MA 02139.

10.1021/jf0606656 CCC: $33.50
Published on Web 07/07/2006

much in vogue at that time. This approach failed probably because of the restricted database that, for example, contained only one sweet compound. Now, following syntheses over some years we have revisited the problem of developing a reliable SAR for the disubstituted phenylsulfamate class of compounds, and in this work a successful structure−taste relationship is reported for these using a larger database of 70 compounds, a test set of 12, and the classification and regression tree (CART) approach with the 9 predictors recently used (*8*), namely, those obtained from (CPK) models, calculated from the PC SPARTAN PRO program, and Hammett σ values from the literature.

This much larger database (70 compounds), which includes a greater selection and mix of substituents X and Y coupled with the employment of new and more appropriate parameters, and the use of CART analysis, which has already been proven to work well in obtaining a SAR for monosubstituted phenylsulfamates, led to the development of the first, reliable SAR for disubstituted phenylsulfamates. The reliability of the method is shown by the fact that 10 compounds from a test set of 12 are correctly classified in this approach.

MATERIALS AND METHODS

Chemistry. The 42 disubstituted aromatic sulfamates synthesized in this work were made according to literature procedures (*10*−*12*). Chlorosulfonic acid was added dropwise with the aid a pressure-equalizing dropping funnel to a 10-fold excess of dry pyridine stirring in a three-neck round-bottom flask. The flask was also equipped with a calcium chloride drying tube, to ensure anhydrous conditions, and a mercury thermometer. The reaction flask was immersed in an ice−salt−acetone bath and the temperature maintained within a range of −5 to 0 °C. The reaction mixture was allowed to stir for 1 h once the addition of chlorosulfonic acid was complete. An equimolar quantity of the liquid/solid amine (equimolar with chlorosulfonic acid) was dissolved in the minimum volume of dry pyridine and added to the pyridine sulfur trioxide adduct formed in situ. The ice−salt−acetone bath was removed and the reaction mixture allowed to come to room temperature and stirred overnight. Ten percent sodium hydroxide was then added to the solution, and each sulfamate was isolated as its sodium salt. Addition of sodium hydroxide continued with stirring until a pH of ≈10 was achieved, the N−S bond being stable under these alkaline conditions. The solution was left to stir for an additional 20 min, and its pH was monitored. Unreacted amine was separated at this stage by several ether extractions. The aqueous pyridine was removed under reduced pressure on a rotary evaporator to yield the crude sulfamate. The sodium sulfamate salt was purified by repeated recrystallization, generally three or four times, from aqueous ethanol. The percentage of aqueous ethanol used varies according to the sulfamate being purified, and generally it is ≈95%. The pure sulfamate product was filtered and air-dried on a Büchner funnel, and the material was dried further over phosphorus pentoxide in a vacuum desiccator for at least 48 h before elemental analysis was carried out.

In the previous procedure the sulfamating agent pyridine sulfur trioxide adduct was formed in situ by the reaction of pyridine with chlorosulfonic acid at a temperature below 0 °C. During the workup process sodium hydroxide is introduced, producing the desired sodium sulfamate salt as well as the undesirable impurity sodium chloride, which was removed by recrystallization. Purification, however, of some sulfamates was difficult due to the presence of sodium chloride, which proved troublesome to remove. In an alternative procedure solid pyridine sulfur trioxide adduct was introduced directly into the reaction rather than its being formed in situ. This obviates the use of chlorosulfonic acid and hence prevents the formation of sodium chloride.

Commercially available pyridine sulfur trioxide complex was allowed to equilibrate in excess dry pyridine at room temperature (*11*, *12*). An equimolar quantity of amine was first dissolved in 30 mL of dry pyridine, added under anhydrous conditions, and allowed to stir overnight at room temperature. The sulfur trioxide adduct is hygroscopic, so to minimize the introduction of moisture into the reaction system, the amine was administered completely in one addition. Ten percent sodium hydroxide was then added to the solution with stirring until a pH of ≈10 was achieved. The solution was left to stir for an additional 20 min. The aqueous layer was washed several times with ether to remove any unreacted organic amine that may have been present. The aqueous pyridine solvent was removed on a rotary evaporator. The resulting crude sodium sulfamate was extracted and recrystallized several times from hot ≈95% ethanol before it was filtered and dried in a desiccator.

*(i) Synthesis of Compound **81**.* The sulfamation of methyl 3-amino-4-methylbenzoate was achieved using the above procedure with some modifications. The sulfamation process was initially carried out as described previously; however, sodium carbonate was utilized, producing compound **81** {sodium [5-(methoxycarbonyl)-2-methylphenyl]-sulfamate}, instead of sodium hydroxide as it is a milder base and is not as likely to cleave the ester group as sodium hydroxide tends to do, thereby producing the disodium 4-methyl-3-(sulfonatoamino)-benzoate salt (*13*). The solvent was removed as before on a rotary evaporator and the crude sulfamate purified by repeated recrystallization from aqueous ethanol. The resulting sodium sulfamate was dried in a desiccator for several days before it was subjected to elemental analysis.

(ii) Synthesis Using 2-Methylpyridine (for Anilines Containing One or Two Electron-Withdrawing Substituents). A number of primary aromatic amines possessed substituents that are strongly electron-withdrawing, for example, NO_2, and their sulfamation could not be achieved under the conditions described earlier. This was overcome by using the following method.

A three-neck 250 mL round-bottom flask, which was charged with a 10-fold excess of dry 2-methylpyridine (α-picoline), was fitted with a stoppered pressure-equalizing dropping funnel, a thermometer, and a calcium chloride drying tube (*12*). The round-bottom flask was in turn cooled within a temperature range of −5 to 0 °C using an acetone−ice−salt bath in a Dewar flask. The pressure-equalizing dropping funnel contained chlorosulfonic acid, which was added dropwise to ensure that the temperature did not exceed 0 °C, reacted with the 2-methylpyridine to give the 2-methylpyridine sulfur trioxide adduct, which formed in situ. Upon completion of the acid addition, the dropping funnel was rinsed with dry 2-methylpyridine and the washings were allowed to enter the reaction vessel. The cooled solution was allowed to stir for an additional 20 min before the acetone−ice−salt bath was removed. The round-bottom flask was heated with stirring to within a temperature range of 70−80 °C using an oil bath. An equimolar quantity of the liquid/solid amine (equimolar with chlorosulfonic acid) was dissolved in dry 2-methylpyridine and added directly to the 2-methylpyridine sulfur trioxide solution. The 2-methylpyridine sulfur trioxide, like the pyridine sulfur trioxide adduct mentioned previously, is quite hygroscopic, so the target amine was introduced completely in a single addition so as to minimize the reaction system's exposure to moisture. The heat source was removed after 1 h, and the sulfamation reaction was allowed to continue with stirring at room temperature overnight. As before, 10% sodium hydroxide was added to the solution until a pH of ≈10 was achieved. The alkaline solution was stirred for an additional 20 min, and its pH was reassessed to ensure the correct range remained before ether extraction was begun. Unreacted amine was removed from the aqueous layer with several ether extractions. The aqueous 2-methylpyridine solvent was removed under reduced pressure, yielding a crude solid material. The resulting impure sodium sulfamate product was recrystallized several times from aqueous ethanol as described previously in the initial procedure. The pure product was finally filtered and dried in a desiccator.

*(iii) Synthesis of Compound **80**.* The sulfamation of 2-amino-4-nitrophenol to **80** [disodium (5-nitro-2-oxidophenyl) sulfamate] could not be accomplished using conventional techniques as sulfation of the hydroxyl and sulfamation of the amino groups are competing reactions. However, sulfamation of the aminophenol was achieved using procedures based on those of Boyland and Manson (*14*, *15*) as follows.

A three-neck 250 mL round-bottom flask was charged with a 10-fold excess of 2-methylpyridine (α-picoline), fitted with a pressure-equalizing dropping funnel and a thermometer, while anhydrous conditions were ensured once again with the use of a calcium chloride drying tube (*10*−*12*). The 2-methylpyridine sulfur trioxide adduct was

Table 1. Concentration and pH Values of the Four Primary Standards

taste	standard	concentration (M)	pH
sweet	sucrose	4.4×10^{-2} (1.5%)	5.05
sour	citric acid	5.2×10^{-4} (0.01%)	3.69
bitter	quinine sulfate	6.4×10^{-6} (0.0005%)	5.03
salt	sodium chloride	3.4×10^{-2} (0.2%)	5.60

formed in situ with the dropwise addition of chlorosulfonic acid while ensuring the solvent temperature did not exceed 0 °C. The round-bottom flask was cooled using an acetone−ice−salt bath. With completion of chlorosulfonic acid addition the cooled solution was allowed to stir for an additional 20 min before the acetone−ice−salt bath was removed. The round-bottom flask was heated with stirring to within a temperature range of 70−80 °C using an oil bath. An equimolar quantity of 2-amino-4-nitrophenol was added to the 2-methylpyridine solution with stirring. The heat source was removed after 1 h, and the sulfamation reaction was allowed to stir overnight at room temperature under anhydrous conditions. To the resulting 2-methylpyridine solution was added 10% sodium hydroxide, resulting in an alkaline mixture; the pH of this mixture was adjusted to a value of 6 using concentrated hydrochloric acid, and this pH change was monitored using a pH-meter (*14*, *15*). The resulting acidic mixture was extracted several times with diethyl ether, after which it was made alkaline with further additions of sodium hydroxide. The alkaline solution was stirred for an additional 20 min and its pH assessed. A solid mass resulted after the aqueous 2-methylpyridine solvent was removed under reduced pressure. Finally, the crude product **80** was purified by several recrystallizations from aqueous ethanol in a similar fashion as described previously before it was dried over phosphorus pentoxide under reduced pressure in a desiccator.

Characterization of Disubstituted Phenylsulfamates. All 42 disubstituted phenylsulfamates gave C, H, and N microanalysis within ±0.5% except compounds **65**, **66**, **71**, **78**, and **82**. Their analyses are as follows: compound **65**, theoretical C, 33.07%; H, 2.75%; N, 11.02%; found C, 33.31%; H, 3.52%; N, 11.34%; compound **66**·$1H_2O$, theoretical C, 34.40%; H, 4.14%; N, 4.46%; found C, 34.79%; H, 4.72%; N 4.66%; compound **71**·$1H_2O$, theoretical C, 42.35%; H, 5.53%; N, 5.49%; found C, 42.06%; H, 5.31%; N, 6.08%; compound **78**·$2H_2O$, theoretical C, 28.97%; H, 3.82%; N, 9.65%; found C, 28.46%; H, 3.19%; N, 9.51%; compound **82**·$^2/_3H_2O$, theoretical C, 35.51%; H, 3.87%; N, 4.14%; found C, 35.40%; H, 3.70%; N, 3.54%. In the first group of 40 compounds synthesized previously the C, H, and N microanalysis of four compounds also deviated slightly (*9*). Each disubstituted phenylsulfamate was characterized by ^{1}H and ^{13}C NMR and IR spectroscopy. The expected peaks were observed in both the proton and carbon-13 spectra. Each sulfamate gave the characteristic IR frequencies, νNH 3400−3190 cm^{-1}, νNS 730−660 cm^{-1}, νSO_3 (asymm) 1070−1040 cm^{-1}, νSO_3 (symm) 1203−1170 cm^{-1}, and νSO_3 (asymm) 1240−1210 cm^{-1}. C, H, and N microanalysis indicated that some of the products crystallized with various quantities of water of hydration. Each product was tested for sulfate and sulfamate. Recrystallization continued until the sulfate test was negative and a clean sharp sulfamate test resulted.

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 FT-IR spectrum 1000, whereas a Perkin-Elmer 2400 series II analyzer was utilized for C, H, and N elemental microanalysis. pH determinations for all sodium sulfamate salts were made using a Jenway model 3310 pH-meter buffered at pH 4.0, 7.0, and 9.2.

Sensory Analysis of Disubstituted Phenylsulfamates. A panel of subjects was used to determine the taste portfolio of the 42 disubstituted phenylsulfamates synthesized in this work, utilizing a sip and spit methodology (*8*, *9*). Taste appraisals of sulfamates synthesized previously (*9*), compounds **1**−**40**, were evaluated by a panel consisting of eight assessors, for the compounds in **Table 2** generally five assessors were used unless indicated in the footnotes, and for the compounds in the test set, namely, **51**, **52**, **54**, **71**−**73**, **75**, **76**, **78**−**80**, and **82**, five assessors were used. All disubstituted phenylsulfamates were tasted as 0.01 M solutions, a concentration that is found to give detectable tastes. They were prepared using distilled water and stored for a short time in grade B volumetric flasks, and taste evaluation was carried out at 18 ± 0.5 °C within 24 h of their preparation. Each sulfamate solution was presented to an experienced panel of subjects in clean white plastic cups. Panelists were subjected to no more than five samples per tasting session, and fixed aliquots of 8 mL were used.

To ensure the panelists were capable of describing the taste(s) characteristic of a sulfamate solution, a period of training was necessary. Training consisted of subjecting each panelist to a series consisting of four primary standards, sweet, bitter, sour, and salty. The concentration and pH of each primary standard are given in **Table 1** (*16*, *17*). Each of the standards was tasted at or above their recognition threshold to enable the panelists to detect a definite taste. It is widely acknowledged that distilled water possesses a mild sweet taste; therefore, as a control each panelist also received an 8 mL sample of pure distilled water to ensure confidence in their ability to identify the taste of a sweet solution from that of a bland/tasteless one. Once it was felt that the panelists were competent in their ability to categorize each of the primary standards correctly, they were subjected to the 0.01 M sulfamate test solutions. Each panelist was asked to identify one or more of the primary standards that best described the taste profile of each sulfamate solution under investigation. They were also asked to describe aftertastes, if any were present. An aftertaste may be defined as a delayed taste that lingers in the mouth (*18*). This delayed taste appears within 3−5 s after the initial taste.

Each of the 42 disubstituted phenylsulfamates synthesized in this study was tasted once, and their taste portfolios are recorded in the tables. Previous work with sodium arylsulfamates has shown that multiple taste tests on three compounds (one S, one N, and one N/S) with five assessors gave an average deviation of ±9%. In that study each compound was presented five times to each assessor so that 25 tastings were made on each of the three compounds.

In this work, as in a previous study (*8*), certain rules were used: (i) Percent tasteless was not considered while it was determined whether a compound was predominately sweet (S) or nonsweet (N) and sweet taste detected either initially or as an aftertaste has been grossed into the calculation of the "sweetness value" below. It should be noted also that the percent values given in the headings of **Tables 2** and **5** are not strictly percents, and thus very often they add up to >100; for example, for compound **41** in **Table 2** the total is 200. The reason for this is that there were five assessors and for each taste detected a number 20 (100/5) is assigned. Thus, for **42** two assessors found it to be sour, giving a score of $2 \times 20 = 40$, three assessors found it to be bitter, giving a score of $3 \times 20 = 60$, and all five assessors reported a sweet aftertaste, giving a score of $5 \times 20 = 100$. In places in **Table 2** numbers such as 62.5 and 37.5 (for compound **50**) and 14.3 and 57.1 (for compound **53**) arise due to the use of panels containing six and seven assessors, respectively. For these panels each taste detected by a panelist merits 100/8, that is, 12.5, and 100/7, that is, 14.3, respectively, and the numbers 62.5, 37.5, and 57.1 are simple multiples of these.

It is possible that the taste portfolio of a sulfamate may include substantial amounts of both S and N tastes, and such a hybrid is denoted N/S. Thus, for example, compound **42** in **Table 2** is 60% sweet (summing S and sweet aftertaste) and 100% nonsweet (N, summing sour and bitter) and is therefore labeled N; compound **41** is 100% sweet (S, sweet aftertaste) and 100% nonsweet (N, summing sour and bitter) and is labeled N/S, whereas compound **77** is 80% sweet (S and sweet aftertaste) and 40% nonsweet (sour and bitter) and is labeled S accordingly. Assignment to the N/S category was determined by examination of the percentage difference between total S and total N, if the difference is ≤25%, then the compound is designated N/S; however, if the difference is greater than this, then it is assigned to the N or S category depending on the dominant taste, and each of the sulfamates was assigned to one of the three categories in this way. The 40 disubstituted phenylsulfamates (*9*) synthesized previously were also characterized as N, N/S, or S in a similar fashion during the course of this study. Finally, to aid further analysis of the disubstituted phenylsulfamate set a sweetness value was determined (*8*). This value is a measure of the sweet content of each sulfamate and is based on a scale ranging from 0 (exclusively nonsweet) to 100 (exclusively sweet). This sweetness value is calculated as follows: sweetness value = (% S + % S aftertaste) × 100/total taste (excluding tasteless) % (*8*). For

Table 2. Percentage of Assessors Giving the Tastes of Disubstituted Phenylsulfamates[a]

entry	sulfamate	pH	% sweet	% sour	% bitter	% salty	% tasteless	% sweet aftertaste	sweetness value	predominant taste N, S, N/S
41	2,4-diOMe	6.80	0	40	60	0	0	100	50.00	N/S
42	2-OMe, 5-Me	7.30	0	60	40	0	0	60	37.50	N
43	2-OMe, 6-Me	5.60	0	0	100	0	0	100	50.00	N/S
44	2-Me, 4-OMe	7.61	0	0	100	0	0	60	37.50	N
45	2-Cl, 4-Me	6.90	0	80	20	0	0	20	16.66	N
46	2-Me, 3-Cl	6.90	0	60	40	0	0	80	44.44	N/S
47	2-Me, 4-Cl	7.24	0	0	100	0	0	80	44.44	N/S
48	2-Br, 4-Me	8.90	0	20	80	20	0	80	40.00	N/S
49	2-Me, 4-Br	4.95	0	40	60	0	0	80	44.44	N/S
50	3-Me, 4-Br[b]	6.70	0	62.5	37.5	0	0	100	50.00	N/S
74	2,3-$(CH_2)_3$	7.44	40	0	0	20	80	80	82.71	S
77	2-Et, 4-Br	10.27	20	20	20	0	40	60	66.66	S
53	3-NO_2, 4-Cl[c]	10.59	14.3	57.1	57.1	0	0	28.6	27.31	N
81	2-Me, 5-CO_2Me	5.44	0	0	40	0	60	100	62.50	S
55	3-NO_2, 4-Me[b]	3.80	0	50	50	0	0	100	50.00	N/S
56	3-F, 4-Me	7.10	0	40	60	0	0	60	37.50	N
57	2-Cl, 4-F[d]	5.81	0	33	83	0	0	66	36.26	N
58	2-F, 4-Cl	6.90	0	20	60	0	0	80	50.00	N/S
59	2-Cl, 4-Br[c]	–	0	28.6	85.7	0	0	71.4	38.44	N
60	2-F, 4-Br	7.24	0	40	60	0	0	40	28.57	N
61	2-F, 4-I[d]	6.33	0	67	17	0	17	33	28.80	N
62	2,4-diBr[c]	10.87	0	29	86	0	0	0	0.00	N
63	2-OMe, 5-Cl	6.40	0	40	60	0	0	80	44.44	N/S
64	2-NO_2, 4-OMe[c]	3.87	0	14	86	0	0	100	50.00	N/S
65	2-NO_2, 5-Me	2.30	0	100	0	0	0	0	0.00	N
66	2-Br, 4-iso-Pr	3.10	0	0	100	0	0	0	0.00	N
67	3-I, 4-Me	11.81	0	0	100	0	0	0	0.00	N
68	3,4-OCH_2CH_2O	11.60	10	20	50	20	0	0	10.00	N
69	3,4-$(CH_2)_3$	9.70	10	10	40	10	30	0	14.28	N
70	3-Cl, 4-C_4H_8NO	9.75	0	100	0	0	0	0	0	N

[a] All compounds were tasted as 0.01 M solutions made in distilled water of pH varying from 5.7 to 5.9. The solutions were tasted by the assessors within 24 h of preparation. [b] Eight assessors were used. [c] Seven assessors were used. [d] Six assessors were used.

compounds **42**, **41**, and **77** the calculations are as follows: **42**, 60-(100)/160 = 37.50; **41**, 100(100)/200 = 50.00; and **77**, 80(100)/120 = 66.66.

RESULTS AND DISCUSSION

In attempting to develop a reliable structure–taste relationship with good predictive ability, we decided to forego the use of linear or quadratic discriminant analysis, which had not proved so successful in analyzing the monosubstituted phenylsulfamates, and instead to try again the CART analysis, which produced a tree that correctly classified 61 compounds from a training set of 75 and correctly predicted the tastes of 6 from a test set of 8 monosubstituted phenylsulfamates (*8*). The same nine predictors, namely, the CPK-derived length (*x*), height (*y*), width (*z*), and volume (V_{CPK}) of the XYC_6H_3- part of XYC_6H_3-$NHSO_3Na$, were measured, the PC SPARTAN PRO '02 (Wave Function Inc.) program gave HOMO, LUMO, E_{solv}, and another volume $V_{Spartan}$, and, finally, Hammett σ values were summed for the X and Y substituents using the values from the literature in **Table 3**. The equilibrium geometry for each molecule was obtained using the empirical AM1 module within the program. For the calculation of E_{solv} the SM 5.4 procedure within the program was utilized.

Five of the predictors, namely, *x*, *y*, *z*, V_{CPK}, and $V_{Spartan}$, that are used are various measures of size and volume, and we have used these because the bulkiness of the R moiety in $RNHSO_3$-Na has been shown to be important in developing structure–taste relationships for the sulfamates (*7*, *19*, *20*). The electronic properties HOMO and LUMO show electron-rich areas such as lone pairs and regions susceptible to nucleophilic attack, respectively. E_{solv} is the sum of the aqueous solvation energy and the gas-phase total energy of each molecule: $E_{solv} = E_{aq.solvation} + E_{total}$. The aqueous solvation term should give an indication of hydrophobic/hydrophilic interactions for each compound. Because our compounds are highly soluble sodium salts, it has not been possible to measure Hansch π parameters. The calculation of the five Spartan-derived properties was based on the whole anion, $RNHSO_3^-$ and a charge of −1 was applied to all of the sulfamates, except in the case where X = O^- (compound **80**) when −2 was used. The Hammett σ values are measures of inductive and resonance effects operating from the ortho-, meta-, and para-positions, respectively, of the aryl ring. To calculate a combined σ effect, $\sum\sigma$, the algebraic addition of the two σ values for X and Y was summed in each case. Some

Table 3. Hammett Sigma Values[a] Used in the Calculation of $\Sigma\sigma$

substituent	$\sigma_{o-(2,6)}$	$\sigma_{m-(3,5)}$	$\sigma_{p-(4)}$
F	0.54	0.34	0.15
Cl	0.67	0.37	0.23
Br	0.70	0.39	0.23
I		0.35	0.24
Me	–0.15	–0.06	–0.14
Et	–0.13		
iso-Pr	–0.15		–0.10
sec-Bu	–0.18		
OMe		0.12	–0.21
OEt			–0.25
CF_3		0.44	
NO_2	1.40		0.78
CO_2Me		0.33	
O^-	–1.10		
sec-Bu	–0.28		
C_4H_8NO			–0.82

[a] Hammett σ values were obtained from: Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.

Table 4. Parameters, Predominant Taste, and Sweetness Value for Disubstituted Phenylsulfamates

entry	sulfamate	x^c	y^d	z^e	$V_{CPK}{}^f$	$V_{Spartan}{}^g$	HOMO^h	LUMO^i	$E_{solv}{}^j$	$\sum\sigma^k$	taste	sweetness value
1	2,3-diF[a]	6.19	3.37	6.90	143.92	183.66	–5.4866	3.0219	–302.548	0.88	N	0.00
2	2,4-diF[a]	6.73	3.37	6.85	155.43	183.66	–5.3210	3.0384	–302.470	0.69	N	0.00
3	2,5-diF[a]	6.16	3.37	7.44	154.36	183.69	–5.4665	3.0217	–303.843	0.88	N/S	50.00
4	2,6-diF[a]	6.18	3.38	7.39	154.13	183.80	–5.2950	2.9937	–301.880	1.08	N	0.00
5	3,4-diF[a]	6.29	3.37	6.89	145.87	183.96	–5.4385	3.0394	–300.921	0.49	N	0.00
6	3,5-diF[a]	5.74	3.38	7.42	144.04	183.88	–5.6416	3.0475	–302.343	0.68	S	88.88
7	2,3-diCl[a]	6.69	3.48	7.60	176.64	205.00	–5.4763	3.0087	–226.658	1.04	N	11.11
8	2,4-diCl[a]	8.19	3.48	7.70	219.11	205.59	–5.4225	3.0724	–227.158	0.90	N	0.00
9	2,5-diCl[a]	6.83	3.48	9.07	215.07	205.72	–5.4970	2.9824	–227.747	1.04	N	0.00
10	3,4-diCl[a]	7.09	3.46	7.61	186.25	205.82	–5.4947	3.0349	–226.666	0.60	N	37.50
11	3,5-diCl[a]	5.69	3.48	9.01	178.29	206.39	–5.6097	2.9779	–227.258	0.74	N/S	50.00
12	2,3-diMe[a]	6.90	3.93	7.33	199.18	209.34	–5.0555	3.5681	–228.001	–0.21	N/S	46.66
13	2,4-diMe[a]	7.62	3.74	7.42	211.46	210.22	–4.9894	3.6356	–229.269	–0.29	N	38.46
14	2,5-diMe[a]	6.90	3.93	8.28	225.03	210.17	–5.0585	3.5676	–229.492	–0.21	N/S	50.00
15	2,6-diMe[a]	6.90	3.80	8.42	220.54	209.50	–5.4532	3.5046	–221.504	–0.30	N	27.27
16	3,4-diMe[a]	6.53	3.85	7.22	181.33	210.20	–5.0384	3.6108	–229.918	–0.20	N/S	42.85
17	3,5-diMe[a]	5.66	3.86	8.51	185.79	210.85	–5.1155	3.5807	–230.593	–0.12	N/S	50.00
18	2,4-diNO_2[a]	7.13	4.83	7.56	260.53	229.98	–6.7000	1.5447	–206.432	2.18	N	0.00
19	3,4-diOMe[a]	7.75	5.79	8.58	253.40	232.88	–5.1959	3.4316	–289.105	–0.09	N/S	43.75
20	3,4-OCH_2O[a]	6.92	3.78	6.83	178.90	200.79	–5.1308	3.3063	–276.422	–0.09	N	20.00
21	3-Cl, 4-F[a]	6.35	3.48	7.64	168.56	195.12	–5.4311	3.0012	–263.058	0.52	N	0.00
22	2-F, 5-Me[a]	6.17	3.76	8.03	186.16	197.10	–5.2201	3.2951	–268.003	0.48	N	0.00
23	2-Me, 5-F[a]	6.98	3.76	7.79	204.50	196.66	–5.3091	3.3040	–266.334	0.19	N	9.09
24	2-Me, 3-F[a]	6.68	3.78	7.56	190.69	196.28	–5.3079	3.3208	–264.784	0.19	N	33.33
25	3-Cl, 4-Me[a]	6.53	3.81	7.62	189.64	208.13	–5.2875	3.3012	–228.356	0.23	N	38.46
26	2-Me, 5-Cl[a]	6.83	3.81	8.74	227.45	207.88	–5.2998	3.2600	–228.395	0.22	N	0.00
27	2-F, 5-NO_2 [a]	6.22	4.83	8.00	240.27	207.34	–5.9041	1.7269	–254.474	1.25	N/S	43.75
28	2-Me, 5-NO_2 [a]	6.93	4.83	8.60	288.17	220.41	–5.7422	1.9507	–218.152	0.56	N/S	50.00
29	2-Cl, 5-NO_2 [a]	6.98	4.83	8.92	300.67	218.14	–5.9226	1.7339	–215.843	1.38	N/S	50.00
30	2-OMe, 5-NO_2 [a]	6.79	5.64	9.26	354.15	232.41	–5.8494	1.8204	–247.481	1.05	N	0.00
31	2-OMe, 5-Cl[a]	6.72	4.84	9.35	303.72	219.98	–5.4120	3.0810	–258.864	0.71	N	38.88
32	2-NO_2, 4-OEt[a]	8.23	5.14	7.53	318.58	251.75	–5.8906	2.0750	–255.277	1.15	N	11.11
33	2-OMe, 4-NO_2 [a]	8.10	4.40	8.20	292.39	232.62	–6.0113	1.9312	–249.458	1.12	N	11.11
34	2-F, 5-CF_3 [a]	6.37	4.75	8.46	255.89	217.15	–5.6790	2.5436	–413.189	0.98	N	0.00
35	2-Cl, 5-CF_3 [a]	7.15	4.62	9.42	311.40	227.90	–5.7060	2.5494	–374.600	1.11	N	35.71
36	2-Br, 5-CF_3 [a]	7.49	4.80	9.60	345.05	234.16	–5.7612	2.4900	–362.447	1.14	N	0.00
37	2-Me, 6-iso-Pr[a]	7.86	4.92	9.70	347.27	247.70	–5.3378	3.5067	–225.357	-0.30	N	37.50
38	2-Et, 6-iso-Pr[a]	8.00	4.64	10.62	394.21	268.57	–5.4027	3.4810	–229.463	-0.28	N	0.00
39	2-Et, 6-sec-Bu[a]	7.68	6.55	11.18	562.37	288.76	–5.4849	3.4664	–234.695	-0.41	N	0.00
40	3-NO_2, 4-F[a]	6.28	4.79	7.57	227.75	206.86	–5.8996	1.6113	–251.657	0.86	N	0.00
41	2,4-diOMe[b]	9.05	5.24	7.95	377.00	234.10	–5.2106	3.2751	–289.612	0.13	N/S	50.00
42	2-OMe, 5-Me[b]	6.89	4.59	9.17	290.00	222.22	–5.1772	3.3843	–260.405	0.28	N	37.50
43	2-OMe, 6-Me[b]	6.93	4.88	9.19	310.79	220.27	–5.3954	3.3751	–252.324	0.19	N/S	50.00
44	2-Me, 4-OMe[b]	8.95	4.81	7.44	320.29	222.08	–5.0998	3.4567	–259.165	-0.36	N	37.50
45	2-Cl, 4-Me[b]	7.57	3.51	7.61	202.20	208.03	–5.1954	3.3552	–229.245	0.53	N	16.66
46	2-Me, 3-Cl[b]	7.18	3.52	7.51	189.90	207.06	–5.3014	3.2686	–226.218	0.22	N/S	44.44
47	2-Me, 4-Cl[b]	8.22	3.50	7.46	214.62	207.93	–5.2297	3.3431	–228.077	0.08	N/S	44.44
48	2-Br, 4-Me[b]	7.71	3.75	7.94	229.56	214.20	–5.2554	3.2882	–217.360	0.56	N/S	40.00
49	2-Me, 4-Br[b]	8.67	3.77	7.48	244.49	214.25	–5.3082	3.3247	–217.186	0.08	N/S	44.44
50	3-Me, 4-Br[b]	7.67	3.76	7.44	214.56	214.39	–5.3569	3.2882	–217.503	0.17	N/S	50.00
74	2,3-$(CH_2)_3$ [b,l]	6.38	3.79	8.53	206.26	217.67	–5.1138	3.5726	–226.428	-0.21	S	82.71
77	2-Et, 4-Br[b]	9.22	5.04	7.81	362.92	233.99	–5.3050	3.3229	–221.584	0.10	S	66.66
53	3-NO_2, 4-Cl[b]	7.06	4.81	7.31	248.24	218.77	–5.9428	1.9226	–212.070	0.94	N	27.31
81	2-Me, 5-CO_2Me[b]	7.18	3.82	11.06	303.35	244.24	–5.3843	2.4596	–301.398	0.18	S	62.50
55	3-NO_2, 4-Me[b]	6.65	6.65	7.46	329.90	219.45	–5.7495	1.8169	–217.235	0.54	N/S	50.00
56	3-F, 4-Me[b]	6.71	3.61	6.83	165.44	197.22	–5.2832	3.3578	–266.255	0.20	N	37.50
57	2-Cl, 4-F[b]	7.06	3.45	7.50	182.68	194.43	–5.3483	3.0423	–264.054	0.82	N	36.26
58	2-F, 4-Cl[b]	7.65	3.49	6.79	181.28	194.99	–5.3944	3.0629	–265.639	0.77	N/S	50.00
59	2-Cl, 4-Br[b]	8.15	3.76	7.34	225.93	212.09	–5.5008	3.0669	–216.443	0.90	N	38.44
60	2-F, 4-Br[b]	8.10	3.75	6.79	206.25	201.32	–5.4691	3.0566	–254.703	0.77	N	28.57
61	2-F, 4-I[b]	7.96	4.12	6.57	215.46	209.83	–5.5142	3.0467	–243.852	0.78	N	28.80
62	2,4-diBr[b]	8.63	3.72	7.67	246.23	218.29	–5.5592	3.0027	–204.655	0.93	N	0.00
63	2-OMe, 5-Cl[b]	7.06	4.64	9.31	304.98	219.98	–5.4120	3.0810	–258.864	0.71	N/S	44.44
64	2-NO_2, 4-OMe[b]	7.89	5.75	7.55	342.52	231.62	–5.9041	2.0697	–249.182	1.19	N/S	50.00
65	2-NO_2, 5-Me[b]	7.39	4.26	8.98	282.70	219.66	–5.8839	2.1725	–222.202	1.34	N	0.00
66	2-Br, 4-iso-Pr[b]	9.02	6.30	7.91	449.49	254.74	–5.2959	3.2722	–226.474	0.60	N	0.00
67	3-I, 4-Me[b]	6.40	3.80	8.27	201.13	222.61	–5.3230	3.0565	–205.113	0.21	N	0.00
68	3,4-OCH_2CH_2O[b]	7.42	4.92	7.41	270.51	219.62	–5.1556	3.4010	–284.136	0.09	N	10.00
69	3,4-$(CH_2)_3$ [b,m]	6.53	3.85	7.22	181.51	217.93	–5.0257	3.6112	–226.256	-0.20	N	14.28
70	3-Cl, 4-C_4H_8NO[b]	11.00	6.45	8.40	597.07	281.18	–5.4724	3.0667	–252.436	-0.45	N	0.00

[a] See ref *9*. [b] Present work. [c] *x* (Å) is the length of the XYC_6H_3– group in $XYC_6H_3NHSO_3Na$. [d] *y* (Å), the width of the XYC_6H_3– group, is offset at 90° to and shares the same plane as the *x*-axis. [e] *z* (Å), the height of the XYC_6H_3– group, is perpendicular to the plane of the *x*- and *y*-axes. [f] V_{CPK} (Å^3) is a product of *xyz* and represents the volume occupied by the XYC_6H_3– group. [g] $V_{Spartan}$ is the volume of a sulfamate as calculated by the PC SPARTAN PRO program. [h] The equilibrium geometry for each molecule was obtained using the semiempirical AM1 module within PC SPARTAN PRO. The electronic property HOMO energy (eV), the energy of highest occupied molecular orbital, was calculated within this software program. [i] LUMO energy (eV), the energy of the lowest unoccupied molecular orbital, was calculated within the PC SPARTAN PRO software program. [j] The aqueous solvation energy E_{solv} (kcal mol^{-1}) was calculated using the SM5.4 procedure within PC SPARTAN PRO. E_{solv} is the sum of the aqueous solvation energy and the total energy of each molecule. [k] Hammett $\sum\sigma$ values were obtained from the data in **Table 3**.

Table 5. Percentage of Assessors Giving the Tastes of Disubstituted Phenylsulfamates for the Test Set[a]

entry	sulfamate	pH	% sweet	% sour	% bitter	% salty	% tasteless	% sweet aftertaste	sweetness value	predominant taste N, S, N/S
71	2-Me, 6-Et	5.11	20	20	80	0	20	60	44.44	N/S
72	2,6-diEt	6.83	0	0	100	0	0	0	0.00	N
73	2,6-di-iso-Pr	9.88	0	0	100	0	0	0	0.00	N
51	2-Cl, 4-NO_2	11.57	0	0	100	0	0	0	0.00	N
75	2-Me, 4-F	6.15	0	20	100	0	20	0	0.00	N
76	3-Me, 4-F	5.67	0	20	80	0	20	40	28.57	N
52	2-NO_2, 4-Cl	3.51	0	40	60	0	0	0	0.00	N
78	2-Me, 4-NO_2	3.21	0	0	100	0	0	0	0.00	N
79	2- NO_2, 4-F	3.25	20	0	80	0	0	0	20.00	N
80	2-O^-, 5-NO_2	8.83	0	0	100	0	80	0	0.00	N
54	2-NO_2, 5-Cl	3.55	0	0	100	0	0	0	0.00	N
82	3,5-diCO_2Me	4.12	0	20	100	0	20	0	0.00	N

[a] All compounds were tasted as 0.01 M solutions made in distilled water of pH varying from 5.7 to 5.9. The solutions were tasted by the assessors within 24 h of preparation; five assessors were used.

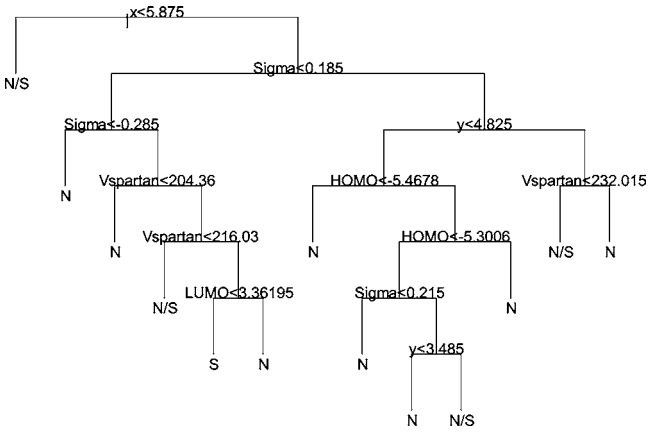

Figure 2. Pruned classification tree using the complete data set of 82 disubstituted phenylsulfamates, 73 of which were correctly classified by the tree (89%).

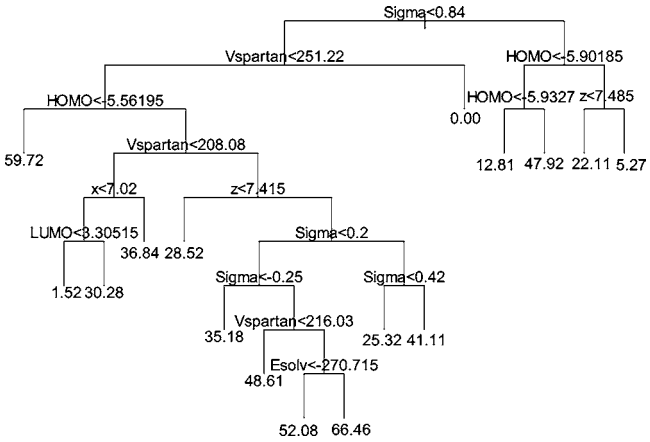

Figure 3. Pruned regression tree using a training set of 70 disubstituted phenylsulfamates, 62 of which were correctly classified by the tree (89%).

Hammett σ values, however, were not always available, so a value for a group similar in structure to the target group was used instead. A Hammett σ value for 2,3-$(CH_2)_3$, **74**, was not given in the literature, so the corresponding Hammett σ value for a 2,3-diCH_3 group was used. Similarly, a Hammett σ value for 3,4-$(CH_2)_3$, **69**, was not present in the literature, so a corresponding Hammett σ value for a 3,4-diCH_3 group was used instead.

The taste data for compounds **1**−**40** have been published (*9*), and the taste data for compounds **41**−**50**, **53**, **55**−**70**, **74**, **77**, and **81** have been given in **Table 2**. For compounds **1**−**40** predominant tastes, that is, N (nonsweet), N/S (nonsweet/sweet), and S (sweet), were derived from the original taste panel data (*9*), and sweetness values or weightings were calculated for them from these data. For the recently synthesized compounds **41**−**50**, **53**, **55**−**70**, **74**, **77**, and **81**, the predominant taste and the sweetness value can be calculated from the data in **Table 2**. The calculation of the sweetness value or weighting was demonstrated earlier for three compounds with various taste portfolios (see Sensory Analysis of Disubstituted Phenylsulfamates). In **Table 4** the values for the nine predictors, the predominant taste, and a sweetness value are given for all 70 compounds that form the training set. Taste data including sweetness values and predominant tastes for the 12 compounds making up the test set are given in **Table 5**.

The sweetness values are very useful when relative sweetness data are not available or cannot easily be measured, as in this instance. The values put all of the tastants on a scale from 0 (N) to 100 (S) with the defined ranges 0−39 for N, 40−60 for N/S, and 61−100 for S; compounds that are particularly nonsweet, for example, **65**−**67** (**Table 2**), are rated 0, and compounds that are strongly sweet, such as **6** (**Table 4**), get a value of 88.88.

Statistical Analysis. The purpose of the analysis is to try to classify all of the known sulfamate tastants of the type $XYC_6H_3NHSO_3Na$ into the three distinct classes N, N/S, and S using the parameters listed above. To this end the computer using random sampling was first asked to remove 12 compounds to act as a test set from the newly synthesized compounds, that is, **41**−**82**, with the proviso that the three S compounds among these would not be removed. This condition was imposed because in the entire data set there are only four S compounds, and they have to be left in the training set to try to obtain a meaningful tree classification containing a sweet node(s). There are, of course, 21 other compounds classified as N/S, and they have been found to have various amounts of sweetness. The computer removed one N/S compound and 11 N compounds. The computer removed compounds **51**, **52**, and **54** (all N) and inserted compounds **74**, **77**, and **81** (all S) in **Table 2**, and this is why the numbering in this table is somewhat randomized.

CART. The use of CART has become popular since the publication of a book (*21*) on the method, and since then there are more than 250 applications of tree methods in the literature (see references in ref *8*). The CART methodology is technically known as binary recursive partitioning. The term "tree" is used because the primary method of display of the results of this statistical analysis is in the form of a binary tree.

Table 6. Parameters, Sweetness Value, and Experimental and Predicted Tastes for Disubstituted Phenylsulfamates Test Set

entry	sulfamate	x^b	y^c	z^d	$V_{CPK}{}^e$	$V_{Spartan}{}^f$	HOMO^g	LUMO^h	$E_{solv}{}^i$	$\sum\sigma^j$	exptl sweetness value	predicted sweetness value	exptl taste	predicted taste
71	2-Me, 6-Et[a]	7.28	4.62	9.18	308.76	228.93	–5.6979	3.4996	–226.050	–0.28	44.44	59.72	N/S	N/S
72	2,6-diEt[a]	7.57	5.22	9.80	387.25	248.45	–5.7941	3.4863	–230.643	–0.26	0.00	59.72	N	N/S
73	2,6-di-iso-Pr[a]	6.86	5.46	10.69	400.40	288.47	–6.0077	3.4426	–234.615	–0.30	0.00	0.00	N	N
51	2-Cl, 4-NO_2 [a]	8.02	4.82	7.46	288.38	218.29	–6.1040	1.8488	–217.676	1.45	0.00	12.81	N	N
75	2-Me, 4-F[a]	7.22	3.78	7.54	205.78	194.99	–5.1270	3.3297	–261.165	0.00	0.00	36.84	N	N
76	3-Me, 4-F[a]	6.23	3.78	7.37	173.56	197.17	–5.1826	3.3324	–266.092	0.09	28.57	30.28	N	N
52	2-NO_2, 4-Cl[a]	7.75	4.83	7.44	278.50	217.42	–6.0106	1.9690	–217.902	1.63	0.00	12.81	N	N
78	2-Me, 4-NO_2 [a]	7.83	3.66	7.53	215.79	219.48	–5.8784	1.9911	–211.846	0.63	0.00	59.72	N	N/S
79	2- NO_2, 4-F[a]	7.44	3.98	7.83	231.86	207.96	–6.0015	1.8062	–253.398	1.55	20.00	12.81	N	N
80	2-O^-, 5-NO_2 [a]	7.06	3.78	8.22	219.37	209.16	–0.9617	6.1191	–348.767	–0.39	0.00	35.18	N	N
54	2-NO_2, 5-Cl[a]	6.01	4.78	8.77	251.94	217.40	–6.1086	1.9229	–218.591	1.77	0.00	12.81	N	N
82	3,5-diCO_2Me[a]	7.58	3.79	11.95	343.30	282.42	–5.7450	2.1108	–379.277	0.66	0.00	0.00	N	N

[a] Present work. [b] For x, y, z, V_{CPK}, $V_{Spartan}$, HOMO, LUMO, E_{solv}, and $\sum\sigma$, see **Table 4** footnotes.

The process is binary because the parent nodes within the tree are always split into two child nodes (partitioning), and the process is recursive because it can be repeated by treating each child node as a parent. At each partition step the program decides which predictor best separates out the residual (unfitted) compounds in terms of, in the present case, N, N/S, or S, or it uses the sweetness values to categorize the members of the set. In this way one can go on until there is only one value at the final node, but this would be overfitting of the taste or other data and the derived tree would be of little use for predictions. Thus, the program provides a tree that has been shortened, known as a "pruned" tree.

The use of CART is attractive to researchers because when a tree is produced, it is easy to interpret, the method is nonparametric, that is, it makes no assumptions about the relationships between the descriptors, and it tends to give high classification rates and good predictability. There is some collinearity between the descriptors used, but an advantage of using regression trees is that they are much less susceptible to multicollinearity problems than normal regression as each descriptor is considered on its own at each step and the problem of singularity does not arise. Recently, CART methods have been used successfully by us to derive tastant SARs for a series of structurally diverse heterosulfamates (*7*) and for monosubstituted phenylsulfamates (*8*).

Basically, the CART procedure is employed to gain a better understanding of the dependence of the response variables (N, N/S, and S or sweetness values) on the structure of the relationships of potential explanatory parameters (predictors) and their combinations together with their high-order interactions. If the response variable is binary, such as N, N/S, S, the procedure produces a classification tree, whereas a regression tree is produced if the response variable is continuous, as with the sweetness values that run from 0 to 90.

Having been reasonably successful with CART analysis in our recent development of a good regression tree for the monosubstituted phenylsulfamates (*8*), we decided to use CART analysis again in seeking an SAR for the disubstituted phenylsulfamates. As in the previous work classification trees were first built using the S-Plus 6.1 (Insight) statistical package. Initially, all 82 compounds in the data set were used to construct some classification trees. The best tree obtained is shown in **Figure 2**, and this classified 73 compounds (11% misclassification) using 6 descriptors and 13 terminal nodes; the percentages of S, N/S, and N misclassified were 50, 19, and 5%, respectively. Other trees generated were less successful, and, for example, using 7 descriptors and 12 terminal nodes, 68 compounds were classified (17% misclassified) and the percentages of S, N/S, and N misclassified were 100, 14, and 12%, respectively. When trees were further pruned to reduce the number of terminal nodes, the important sweet terminal node disappeared and the classification rate deteriorated.

Another classification tree was now derived by omitting the 12 compounds previously randomly selected by the computer and using the remaining 70 as a training set. The best classification tree produced from this "training set of 70" classified 61 compounds (13% misclassified) using 6 descriptors and possessing 11 terminal nodes; the percentages of S, N/S, and N misclassified were 25, 25, and 7%, respectively. However, only 7 of the 12 compounds in the test set were correctly classified (42% misclassified). Further attempts to improve on this again led to either an increase in the number of compounds misclassified and/or a loss of the vital sweet terminal node.

The alternative option was to get the program to construct a regression tree using the sweetness values in **Table 4** for the 70 compounds of the training set and the descriptors given above. The best tree is shown in **Figure 3**, and this was selected using 10-fold cross-validation and employed 7 descriptors (x, z, $V_{Spartan}$, HOMO, LUMO, E_{solv}, and σ) and 16 terminal nodes to classify 62 compounds correctly (11% misclassified); the percentages of S, N/S, and N misclassified were 25, 33, and 0%, respectively. The performance of this regression tree in **Figure 3** can be gauged from the following: (i) it had a Pearson correlation coefficient of 0.87, giving a P value of <0.0001; (ii) it classifies correctly 62 of the 70 compounds in the training set; and (iii) using the descriptor values and the sweetness values for the test set in **Table 6**, it was found to have a high predictive power, classifying 10 of the 12 compounds (17% misclassified) in the test set.

Tree results can be represented in a useful alternative way using a scatterplot of the experimental taste panel data in the form of sweetness values plotted against the predicted sweetness values. Two such plots are used here to examine the tree in **Figure 3**. In **Figure 4** a scatterplot of all 70 compounds in the training set has been made, and the 8 misclassified compounds are numbered using their entry numbers from **Table 4**. In **Figure 4** some compounds are superimposed on others because they have the same coordinates and, thus, somewhat fewer than 70 compounds can be seen in the scatterplot. The S (green diamond), N/S (red square), and N (black circle) indicate the experimental taste categories for the compounds. Thus, the following compounds that lie outside the prescribed "taste areas" are seen to be misclassified: S (**54**) and N/S (**3**, **12**, **16**, **43**, **46**, **47**, and **58**). The S compound **6** and the N compound **31** are deemed to be classified because **6** is borderline and **31** is only very slightly to the left of the N area. In the second scatterplot

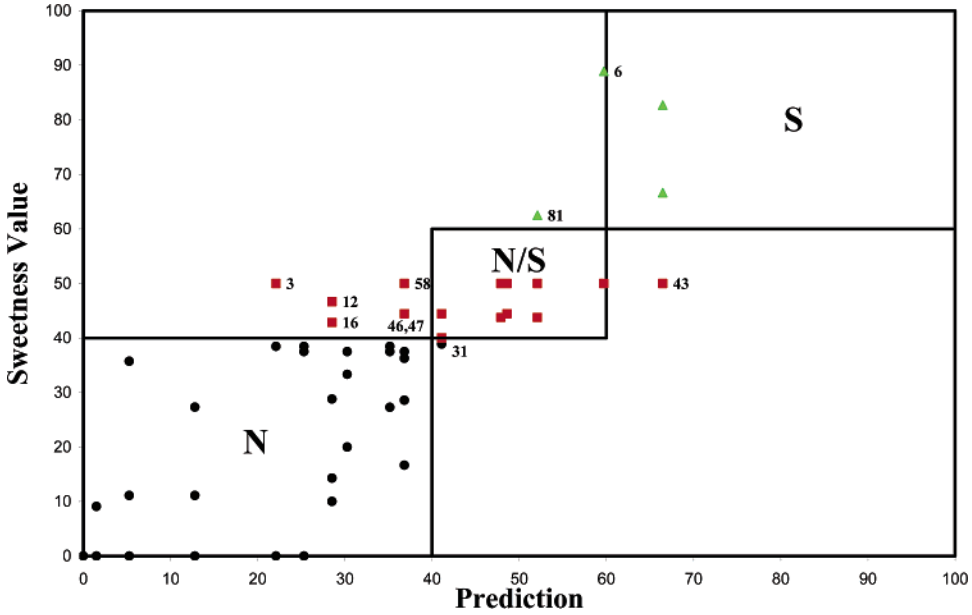

Figure 4. Scatterplot of tree-predicted versus panel sweetness values using the training set of 70 disubstituted phenylsulfamates. Some compounds are superimposed on others because they have the same coordinates, and thus the scatterplot displays fewer than 70 compounds: green diamond, sweet (S); red square, nonsweet/sweet (N/S); black circle, nonsweet (N). The numbered compounds are the ones that are misclassified, although **6** and **31** are deemed to be borderline.

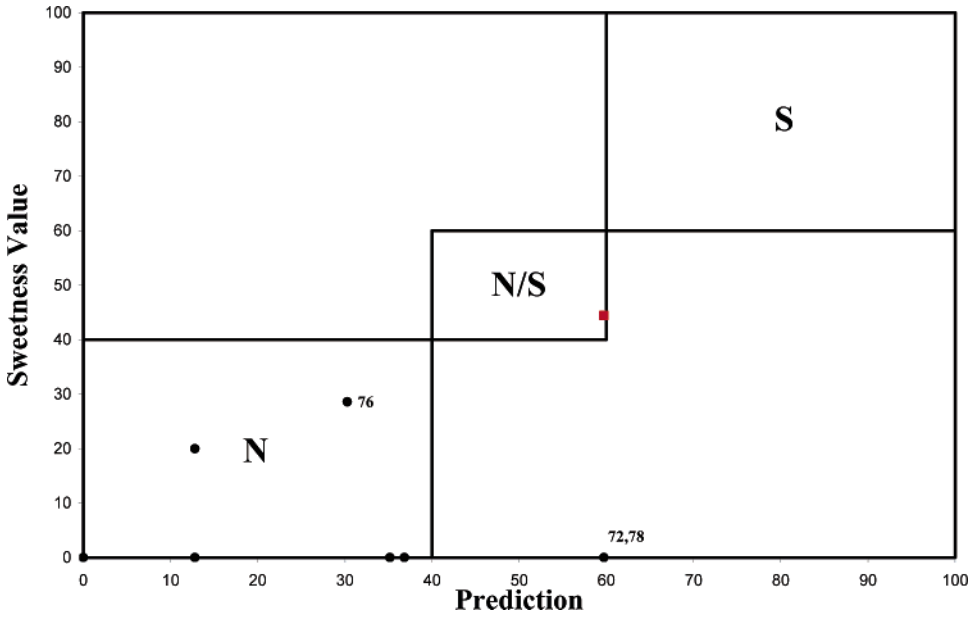

Figure 5. Scatterplot of tree-predicted versus panel sweetness values using the test set of 12 disubstituted phenylsulfamates: green diamond, sweet (S); red square, nonsweet/sweet (N/S); black circle, nonsweet (N). Compounds **72** and **78** are misclassified, and the position of **76** is indicated because it is used in the text to illustrate how the tree works.

(**Figure 5**) results for the test set are graphically displayed, and the two misclassified compounds **72** and **78** are clearly shown. These results are given in **Table 6**.

The use of the tree in **Figure 3** for predictions can be easily illustrated with the following examples: Compound **58** is misclassified by the tree in **Figure 4**, and this can be illustrated as follows using the descriptor values in **Table 4**: its σ (0.77) is <0.84, as required; thus, one follows to the left; at the next node, $V_{Spartan}$ (194.99 for **58**) is <251.22, so one moves left again; now, HOMO should be < -5.56195 but for **58** it is -5.3944, so one moves to the *right*. At the next node $V_{Spartan}$ should be <208.08, and it is, so one moves to the left; x must be <7.02, but **58** has a value of 7.65, so one follows to the *right*, giving a predicted sweetness value of 36.84 compared to the taste panel value of 50.00. Thus, the tree finds **58** to be N but the panel had found it to be N/S, and this is an example of a misclassified compound. In the scatterplot for the test set (**Figure 5**) compound **76**, which was found to be in the N category by the panel, is also found to be N using the tree (**Figure 4**), as follows (using descriptor values in **Table 6**): **76** has a σ of 0.09, which is <0.84 as required by this tree, so one follows to the left; at the next node $V_{Spartan}$ should be <251.22, and **76** has a value of 197.17, so one again follows to the left; then, the next node requires HOMO < -5.56195, and **76** has a value of -5.1826, so one moves *right*; then, $V_{Spartan}$ should be <208.08, and **76** complies with this, so one moves to the left; at the next node x <7.02 is needed, and again the value for **76** is less than that, so a move to the left is made; now, the

requirement is for LUMO <3.30515, and **76** has a value of 3.3324, so one moves *right* to reach a terminal node giving a value of 30.28, which places this compound in the N area in agreement with its experimental sweetness value of 28.57.

In conclusion, the regression tree shown in **Figure 3** with good predictability has been found using CART analysis. Using this tree it should be possible to predict taste with almost a 90% success rate for disubstituted phenylsulfamates. However, it should be noted that its predictive ability is likely to vary for the different tastant categories; it should be very high for nonsweet (N) compounds, have ≈75% reliability for sweet (S) compounds, and have ≈66% accuracy for nonsweet/sweet (N/S) compounds. In simple terms it has a very strong chance of predicting a nonsweet compound, a three of four chance of predicting a sweet compound, and a two of three chance of predicting a nonsweet/sweet compound. Our latest data do suffer from a paucity of sweet compounds, although the new synthesis has resulted in three additional sweet compounds, which we have employed to maximum effect by including them in the training set. From analysis of the pruned regression tree (**Figure 3**) using the three sweet compounds which were correctly classified, that is, **6**, **74**, and **77**, it appears that new sweet compounds will have to have descriptors that fall in the following ranges: either $\sigma < 0.84$, $V_{Spartan} < 251.22$, and HOMO < -5.56 or σ 0.2 to -0.25, $z > 7.4$, $V_{Spartan}$ 216−251, and $E_{solv} > -270$.

The success achieved with the larger database, new parameters, and new approach simply eclipses our previous efforts with disubstituted phenylsulfamates, for which, in fact, we were unable to find an overall SAR (*9*).

ACKNOWLEDGMENT

We acknowledge Dr. J. Newell, Mathematics Department, National University of Ireland, Galway, for contributions to the statistical work. We thank Dottikon, Ethyl Corp., Shell Fine Chemicals, and the University of Reading Chemistry Department, Reading, U.K., for gifts of chemicals.

LITERATURE CITED

(1) Bopp, B. A.; Price, P. In *Alternative Sweeteners*, 3rd ed.; O'Brien Nabors, L. Ed.; Dekker: New York, 2001; Chapter 5, pp 63−85.

(2) Parke, S.; Birch, G. G.; Dijk, R. Some taste molecules and their solution properties. *Chem. Senses* **1999**, *24*, 271−279.

(3) Spillane, W. J.; Morini, G.; Birch, G. G. Determination of apparent molar volumes (AMV) and capacity factors (k_w) for sweet and non-sweet sulfamates, correlation between AMV values, and Corey−Pauling−Koltun volumes and relative sweetness. *Food Chem.* **1992**, *44*, 337−342.

(4) Spillane, W. J.; Hanniffy, G. G. Is the negative charge on $RNHSO_3^-M^+$ an essential requirement for sulfamate sweetness? *J. Agric. Food Chem*. **2003**, *51*, 3056−3059.

(5) Unterhalt, B.; Kerckhoff, L; Mollers, M. Monomethylated 3-thiacyclopentylsulfamic acids and their quality of sweetness. *Sci. Pharm.* **2000**, *68*, 101−108.

(6) Pautet, F.; Daudon, M. Role specifique du groupe aminosulfonate dans I'interactioin des sulfamates a caractere edulcorant avec leur site recepteur. *Pharm. Helv. Acta* **1986**, *61*, 215−217.

(7) 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 (Gainesville, FL, U.S.)* **2003**, *7*, 297−309; available at http://www.arkat-usa.org/ark/journal/2003/107_McKervey/AM-736D/736D.pdf.

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

(9) 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.

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

(11) 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.

(12) Gilbert, E. E. *Sulfonation and Related Reactions*; Wiley-Interscience: New York, 1965; Chapter 7, pp 403−424.

(13) Hopkins, A.; Williams, A. Carboxyl group participation in sulfate and sulfamate group transfer reactions. *J. Org. Chem.* **1982**, *47*, 1745.

(14) Boyland, E.; Manson, D.; Sims, P. The preparation of *o*-aminophenyl. *J. Chem. Soc*. **1953**, 3623−3628.

(15) Boyland, E.; Manson, D. The reaction of chlorosulfonic acid with aminophenols. *J. Chem. Soc.* **1958**, 532−534.

(16) Best, C. H.; Taylor, N. B. *Taste in Physiological Basis of Medical Practices*; Brobeck, J. R., Ed.; Williams and Wilkins: Baltimore, MD, 1973; pp 121−131.

(17) Paulus, K.; Reisch, A. M. The influence of temperature on the threshold values of primary tastes. *Chem. Senses* **1980**, *5*, 11−21.

(18) Horio, T.; Kawamura, Y. Studies on the aftertaste of various taste stimuli in humans. *Chem. Senses* **1990**, *15*, 271−280.

(19) Pautet, F.; Nofre, C. Correlation of chemical structure and taste in the cyclamate series and the steric nature of the chemoreceptor site. *Z. Lebensm. Unters.-Forsch.* **1978**, *166*, 167−170.

(20) Spillane, W. J.; McGlinchey, G. Structure−activity studies on sulfamate sweeteners II. Semiquantitative structure−taste relationship for sulfamate ($RNHSO_3^-$) sweeteners—the role of R. *J. Pharm. Sci*. **1981**, *70*, 933−935.

(21) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth and Brooks: Belmont, CA, 1984.

Received for review March 8, 2006. Revised manuscript received May 31, 2006. Accepted June 5, 2006. The Department of Education, Ireland, Forbairt, Ireland, and British Oxygen are thanked for grants.

JF0606656