CROATICA CHEMICA ACTA CCACAA **75** (2) 475–502 (2002)

ISSN-0011-1643
CCA-2812 *Original Scientific Paper*

A QSPR Study of Sweetness Potency Using the CODESSA Program*

Alan R. Katritzky,[a,**] *Ruslan Petrukhin,*[a] *Subbu Perumal,*[a]
Mati Karelson,[b] *Indra Prakash,*[c] *and Nitin Desai*[c]

[a] *Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P. O. Box 117200, Gainesville, FL 32611–7200, USA*

[b] *Department of Chemistry, University of Tartu, Jakobi St. 2, Tartu 51014, Estonia*

[c] *The NutraSweet Company, 601 E. Kensington Rd., Mt. Prospect, IL 60056–1300, USA*

Received April 23, 2001; revised September 11, 2001; accepted September 12, 2001

A total of 397 natural and artificial comprehensively referenced sweeteners were classified by their structures into nine sets. The sweetness potencies were correlated with quantum chemical and other molecular descriptors using the heuristic and the best multi-linear regression methods of the CODESSA software package. QSPR models (two-parameter unless otherwise indicated) emerged for each subclass of sweeteners with R^2 values of 0.835 for 47 aldoximes, 0.959 for 8 acesulfamates, 0.919 for 9 sulfamates, 0.941 for 10 α-arylsulfonylalkanoic acids, 0.715 for 27 guanidines (0.802 in a three-parameter correlation), 0.769 for 30 ureas/thioureas (0.888 in a three-parameter correlation), 0.905 for 20 natural sweeteners, 0.957 for 7 miscellaneous sweeteners (one-parameter correlation), 0.688 for 87 peptides (five-parameter correlation). A significant global five-parameter QSPR theoretical model with R^2 of 0.686 for the entire set of sweeteners is presented and discussed with reference to the possible existence of single or multiple sweetness receptors.

Key words: QSPR, sweetness, CODESSA, model, prediction, database of sweeteners, correlation.

---

[*] Dedicated to Professor Milan Randić on the occasion on his 70th birthday.

[**] Author to whom correspondence should be addressed. (E-mail: katritzky@chem.ufl.edu)

INTRODUCTION

The development of low-calorie sweeteners for sugar substitution is medically very important, and potentially could solve serious problems of modern human nutrition and health. For instance, they can have a beneficial effect in the management of diabetes and obesity which, in turn, can control a number of other diseases. The important requirements for a sweetener for commercial utility are: (i) high sweetness relative to sucrose which will bring down the cost as well as the caloric value, (ii) early onset of the sweet taste and absence of lingering off-taste, (iii) stability under the conditions of use, (iv) lack of toxicity, and (v) low cost of production. Consequently, the search for sweeteners with an ideal blend of the above features has been continuing since 1898 and many different classes of sweeteners have emerged.

In previous attempts to unravel the factors underlying the sweetness property, many QSPR studies were performed for different sub-sets of sweeteners: dihydrochalcones,[1] perillartines,[2] sulfamates,[3] five membered rings[4] and diverse compounds.[5] Although the above QSPR studies can describe individual classes of sweeteners well, finding successful structure-activity correlations encompassing all classes of sweeteners has remained elusive perhaps unsurprisingly in view of the involvement of multiple receptors.[6] However, significant progress has recently been achieved with the development of the multi-point docking concept[7] and the latest 15-point system[7a] significantly assists the understanding for most known classes of sweeteners. Much of the earlier work has been clearly summarized by van der Heijden:[6c] this author developed geometrical explanations for the sweet taste involving AH···B moieties in the various classes of the sweet substances.

Since 1991, linear multivariable QSPR methodology has been developed by our group within the CODESSA software package.[8a] The CODESSA program is undergoing continuous development and the latest available version has been applied to the sweetness task. The main objective of the present work is to evolve general QSPR models for predicting the degree of sweetness for the global set of sweeteners as well as the different subclasses using the CODESSA software package.[8a] CODESSA has previously been applied to many QSPR relationships.[8c–m]

METHODOLOGY

The Sweeteners' Database

A database of sweeteners was created from literature sources.[1,3,4,6b,7a,10] The general set comprises many subclasses including: aldoximes (perillartines), sulfamates, acesulfamates, (α-arylsulphonyl)cycloalkanoic acids, natural sweeteners including sugars, peptides, guanidines and other substances.

The types of sweeteners in the database and its current extent are depicted in Table I. Currently, 397 structures have been entered into the database; quantitative data for the onset sweetness property were recorded for 351 structures; and for the remaining structures, qualitative data were included as the only information available. In this study, the qualitative assessments for sweetness, bitterness, sourness, and saltiness, are expressed by a range in which 0 means »non-sweet« (for sweetness), 5 means »very sweet« and values between 1 and 4 describe intermediate levels of sweetness, all on a semi-quantitative scale. If a description that a structure is sweet (or bitter, *etc.*) alone is available, a default value of 3 is used. For some structures having both sweet and bitter tastes, additional information about the sweet-bitter blend is given as a ratio in a »comment« field. The current content of the database consisting of the structures, sweetness potencies, literature and other pertinent information is included as Table II in the supplementary material to this article.

TABLE I

Content of the sweeteners' database

Class of sweeteners	Number of structures	
	All	Positive quantitative sweetness
Aldoximes (perillartines)	48	47
Sulfamates	111	9
α-Arylsulphonylcycloalkanoic acids	10	10
Acesulfamates	9	8
Naturally occurring substances	25	20
Peptides	88	87
Guanidines	28	27
Urea and thiourea	39	30
Others	55	7
Compounds belonging to two sub-sets (11) counted twice in the above rows	–16	–7
Total	397	238

Derivation of Descriptors and Regression Analyses

The sweetener database was created using ISIS technology[11] of MDL Information Systems, Inc., and transferred to the CODESSA program. All structures were prepared for MOPAC[12] optimization by the HYPERCHEM program.

The 3D geometry for all compounds have been calculated utilizing (i) the MOLGEO program,[13] (ii) the MM+ molecular mechanics method using HyperChem

TABLE II

The database content

Quantitative				Qualitative			
Sweetness	Bitterness	Sourness	Saltiness	Sweetness	Bitterness	Sourness	Saltiness
351	9	9	9	397	80	33	32

V5.1,[14] (iii) the AM1 parameterizations[15] of MNDO semi-empirical method using the MOPAC V6 software package.[12]

The CODESSA software subsequently produced more than 500 constitutional, topological, geometrical, electrostatic, quantum chemical and thermodynamical molecular descriptors[8b] for each compound and performed the statistical analyses in the descriptor space.

RESULTS AND DISCUSSION

Our aim has been to arrive initially at comprehensive theoretical models for the relative sweetness (RS) of all the sub-sets, with sweeteners having positive quantitative values (Table I), and then to explore the existence of a global model encompassing all known sweeteners. Relative sweetness is defined as the ratio of the concentration of the alternative sweetener to the concentration of sucrose which has an equally sweet taste (dimensionless). The whole set of sweeteners has been classified into nine sub-sets using substructure keys and on the basis of the source of the sweeteners. All the descriptors required for the QSPR study were calculated using the AM1[15] semi-empirical method as realized in the MOPAC[12] computer program. The natural logarithm of the relative sweetness (ln RS) has been used for the QSPR correlations of the present investigation. Both the heuristic (HM)[8a] and the best multi-linear regression methods (BMLR)[8a] have been employed to arrive at the optimum correlations. These methods employ different approaches in the discrete local optimization of the correlation coefficient in the descriptor permutation space. As they are local optimization methods, there is no guarantee that the global optimum on the discrete descriptor space will be found by them. Although both methods use heuristics for eliminating the descriptors with low probability to be involved in the final multiparameter correlation, the approaches to this elimination are completely different.

The best multilinear regression (BMLR) method[8a] searches for the multiparameter regression with the maximum predicting ability using the following strategy:

Step 1. Setting up initial values: (i) the upper limit of the square of the linear correlation coefficient ${R_{\mathrm{min}}}^2$, for two descriptor scales to be considered orthogonal (default value ${R_{\mathrm{min}}}^2 = 0.1$); (ii) the lower limit of the square of the linear correlation coefficient ${R_{\mathrm{nc}}}^2$, for two descriptor scales to be considered non-collinear (default value ${R_{\mathrm{nc}}}^2 = 0.65$); (iii) the maximum number of best correlations to be considered in the search of the best correlations with one descriptor scale added (default value $N_\mathrm{c} = 400$); (iv) the probability level for the Fisher criterion F_{Sel} (default value 0.95).

Step 2. All orthogonal pairs of descriptors i and j (with ${R_{ij}}^2 < {R_{\mathrm{min}}}^2$) are found in a given data set.

Step 3. The property analyzed is treated by using the two-parameter regressions with the pairs of descriptors, obtained in Step 2. The N_c pairs with highest regression correlation coefficients are chosen for performing the higher-order regression treatments.

Step 4. For each descriptor set, obtained in the previous step, a non-collinear descriptor scale, k (with ${R_{ik}}^2 < {R_{\mathrm{nc}}}^2$ and ${R_{kj}}^2 < {R_{\mathrm{nc}}}^2$), is added, and the respective (n+1)–parameter regression treatment is performed. If the Fisher criterion at a 95% probability level, F, is smaller than that for the best n-parameter correlation from previous step, the latter is chosen as the final result and the program proceeds to the output section (Step 5). Otherwise, the N_c descriptor sets with highest regression correlation coefficients are chosen for this step repeated with incremented value of n.

Step 5. The final result, with the maximum value of the Fisher criterion and with the highest cross-validated correlation coefficient is chosen for use in the subsequent prediction part of the program. The results for the best correlation of each rank are saved.

The heuristic method (HM)[8a] for descriptor selection, by contrast, proceeds with a pre-selection of descriptors by eliminating sequentially descriptors which do not match any of the following criteria: (i) Fisher F-criterion greater than one unit; (ii) R^2 value less than a value defined at the start (by default 0.01); (iii) Student's t-criterion less than that defined (by default 0.1); (iv) duplicate descriptors having a higher squared intercorrelation coefficient than a predetermined level (usually 0.8), retaining the descriptor with higher R^2 with reference to the property.

The remaining descriptors are then listed in decreasing order of the correlation coefficients when used in global search for 2-parameter correlations. Each significant 2-parameter correlation by F-criterion is recursively expanded to an n-parameter correlation till the normalized F-criterion remains greater than the startup value. The top N correlations by R^2 as well as F-criterion are saved.

The HM usually produces correlations 2–5 times faster than BMLR, with quality comparable to that obtained with the latter. The results of QSPR studies of the various different classes of sweeteners are discussed in turn below and finally the whole set is considered globally.

Aldoximes

This sub-set consists of 47 aldoximes each containing the group (=N–OH) (Table III). QSPR treatment by CODESSA afforded significant 2- and 3-parameter correlations (Eqs. (1) and (2) respectively).

$$\ln \mathrm{RS} = -(59.8 \pm 4.8)\#\mathrm{O_R} + (0.224 \pm 0.033)\mathrm{SA_C} + (4.18 \pm 0.70) \qquad (1)$$
$$R^2 = 0.835 \quad F = 112 \quad s = 0.099 \quad n = 47$$

(In this and following equations $s = \sqrt{s^2/n}$)

$$\ln \mathrm{RS} = -(51.9 \pm 4.9)\#\mathrm{O_R} - (222 \pm 41)\mathrm{RI^e}_{\mathrm{avg,C}} + (96.2 \pm 14.4)\mathrm{PSA_C} \qquad (2)$$
$$R^2 = 0.857 \quad F = 86.1 \quad s = 0.094 \quad n = 47$$

In these equations, $\#\mathrm{O_R}$ refers to the relative number of oxygen atoms in the sweetener molecule; $\mathrm{SA2_C}$ to the solvent accessible surface area[8b] of the carbon atoms; $\mathrm{RI^e}_{\mathrm{avg,C}}$ to the average electrophilic reactivity index[8b] for a carbon atom; $\mathrm{PSA2_C}$ to the positively charged solvent accessible surface area[8b] of the carbon atoms.

The negative sign of the partition coefficient corresponding to the $\#\mathrm{O_R}$ descriptor, *viz*. the number of oxygens, may indicate competitive interference of oxygen functionalities blocking the oxime groups in the docking process. The surface area related descriptors, $\mathrm{SA_C}$ and $\mathrm{PSA_C}$, have positive signs for their coefficients showing that the larger the area the more the interaction. The descriptor, $\mathrm{RI^e}_{\mathrm{avg,C}}$, defined as sum of squares of the atom orbital contributions in LUMO, involved in the three-parameter equation, may relate to the interactions between the sweetener and the receptor.

Pertinently, no other QSPR study on the sweetness of aldoximes is available for direct comparison with the correlations obtained in the present investigation. Previous work was limited to the development[2a] discriminant functions based on log P and the Randić indexes[8b] ($^1\chi_c$ and $^1\chi_j$) for the sweetness of aldoximes and reported two- and three-parameter relationship (Eqs. (3) and (4), respectively) as the best linear discriminant functions. Po-

TABLE III

Results of correlations for aldoximes (n = 47)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.664	88.7	0.140	0	8.27	0.50	16.6574	Intercept
				1	–63.6	6.8	–9.4204	Relative number of O atoms
2	0.835	112	0.099	0	4.18	0.70	5.9891	Intercept
				1	–59.8	4.8	–12.4166	Relative number of O atoms
				2	0.224	0.033	6.7788	SA-2 of C atoms
3	0.857	86.1	0.094	0	5.121	0.815	6.2805	Intercept
				1	–51.9	4.9	–10.7007	Relative number of O atoms
				2	–222	41	–5.3704	Average electroph. react. index for a C atom
				3	96.2	14.4	6.6578	PSA-2 of C atoms
4	0.871	70.9	0.090	0	–13.7	8.9	–1.5338	Intercept
				1	–46.1	5.4	–8.5431	Relative number of O atoms
				2	–217	40.0	5.4523	Average electroph. react. index for a C atom
				3	75.9	16.9	4.4962	PSA-2 of C atoms
				4	9.27	4.38	2.1174	Maximum bonding contribution of a MO

sitive values of g in Eqs. (3) and (4) indicate a sweet taste while a negative value relates to a bitter taste.

$$g_2(\mathrm{X}) = 0.696 \log P - 0.718\ {}^1\chi_c - 0.220 \times 10^{-5} \tag{3}$$

$$g_3(\mathrm{X}) = 0.595 \log P - 0.773\ {}^1\chi_c - 0.220\ {}^1\chi_j - 0.882 \times 10^{-5} \tag{4}$$

Kier[10a] defined another discriminant function by Eq. (5) with a critical value of –3.27 for y.

$$y = 1.21\ {}^1\chi - 3.88\ {}^4\chi_p \tag{5}$$

The y values greater than the critical value correspond to the sweet taste and those less than it with bitter.

Acesulfamates and Sulfamates

For the acesulfamate sub-set of 8 sweeteners, the best one-parameter correlation (Eq. (6)) is with $CI_{max,C–H,}$ the maximum coulombic interaction for the carbon-hydrogen bond (Table IV), a descriptor which relates to the electrostatic and van der Waals interactions and is correlated with the polarity of the hydrogen-carbon bond which is the closest to oxygen atom in the cycle, and with total electron density on this hydrogen atom, correspondently. The 2-parameter correlation in Table IV also includes descriptors which could influence the free energy of the docking of the sweetener.

$$\ln \mathrm{RS} = -(4.53 \pm 0.38)\mathrm{CI}_{\mathrm{max,C–H}} + (21.7 \pm 1.5) \tag{6}$$

$$R^2 = 0.959 \quad F = 139 \quad s = 0.078 \quad n = 8$$

TABLE IV

Results of correlations for acesulfamates (n = 8)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.959	139	0.078	0	21.7	1.5	14.7937	Intercept
				1	–4.53	0.38	–11.7838	Max coulombic interaction for a C–H bond
2	0.996	677	0.026	0	628	19.8	31.7953	Intercept
				1	–339	11	–31.6183	Max bond order for an O atom
				2	6509	342	19.0291	Average nucleoph. react. index for a N atom

For the sulfamate class of 9 sweeteners, linear regression analysis (Table V) afforded the two-parameter correlation (7).

$$\ln \mathrm{RS} = -(2.49 \pm 0.392)\mathrm{SA_O} + (30.8 \pm 7.71)\mathrm{C'_{MO}} + (26.7 \pm 2.03) \tag{7}$$

$$R^2 = 0.919 \quad F = 34.0 \quad s = 0.058 \quad n = 9$$

The incorporation of the solvent accessible surface area for a oxygen atom (SA_O) and the maximum antibonding contribution into molecular orbital (C'_{MO}) in the above equations can be understood in terms of their possible influence on the interactions associated with the docking complex formation.

In earlier work, a one-parameter relationship (8) between the relative sweetness (RS) of sulfamates and the third order Randić index[8b] ($^3\chi_m$) was reported.[4b]

$$\mathrm{RS} = (16.06 \pm 4.21)\ ^3\chi_m - (10.19 \pm 7.31) \tag{8}$$
$$R^2 = 0.5929 \quad s = 8.41 \quad n = 12$$

Previously, two discriminant functions were developed from geometrical parameters of the structures and the first order Randić index.[3b] A non-linear correlation for sweetness was based on four STERIMOL parameters:[3a]

$$D = [0.25\{(\mathrm{MR} - 24.4/8.77)^2 + (L - 5.68/1.46)^2 + (B_1 - 1.69/0.395)^2 + (B_0 - 3.89/1.15)^2\}]^{0.5} \tag{9}$$
$$F = \frac{D^2}{\mathrm{RSD}^2}, \quad F < F(4.52)_{\alpha\ 0.05}$$

In Eq. (9) L, B_1 and B_0 are STERIMOL parameters while RSD is the residual standard deviation. It is not possible to compare the results of a discriminant analysis with those of multi-linear regression analysis as these two approaches differ in their basic concepts.

TABLE V

Results of correlations for sulfamates (n = 9)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.785	25.6	0.088	0	253	51	–4.9999	Intercept
				1	272	54	5.0597	Minimum valency of a H atom
2	0.919	34.0	0.058	0	26.7	2.03	1.3119	Intercept
				1	2.49	0.392	6.3470	SA-2 of O atoms
				2	30.8	7.71	4.0000	Max antibonding contribution of a MO

A one-parameter correlation between the apparent molar volume (AMV) (which, in turn, correlates well with sweetness) and various types of volume functions calculated by the GEPOL program for sulfamates has been reported by Spillane *et al.*[5a] Another study has reported a correlation between AMV and molecular surface volume (MSV),[5b] but provides no direct relationship between relative sweetness and either the AMV or MSV.

α-*Arylsulfonylalkanoic Acids*

For a series of ten α-arylsulfonylalkanoic acids, the best two-parameter correlation is Eq. (10) (Table VI).

$$\ln \mathrm{RS} = -(30400 \pm 4100)\mathrm{RI}^{\mathrm{e}}{}_{\mathrm{avg,C}} - (2890 \pm 450)\mathrm{RI}^{1\mathrm{e}}{}_{\mathrm{min,S}} + (5.01 \pm 0.11) \quad (10)$$

$$R^2 = 0.941 \quad F = 55.3 \quad s = 0.096 \quad n = 10$$

Eq. (10) includes the minimum electrophilic reactivity index for a carbon atom ($\mathrm{RI}^{\mathrm{e}}{}_{\mathrm{avg,C}}$) and the minimum 1-electron reaction index for a sulphur atom ($\mathrm{RI}^{1\mathrm{e}}{}_{\mathrm{min,S}}$), defined as

$$\mathrm{RI}^{1\mathrm{e}} = \sum_{i \in \mathrm{A}} \sum_{j \in \mathrm{A}} c_{i_{\mathrm{HOMO}}} c_{j_{\mathrm{LUMO}}} / (\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}),$$

where the summations are performed over all atomic orbitals i, j at the given atom, $c_{i_{\mathrm{HOMO}}}$ and $c_{j_{\mathrm{LUMO}}}$ denote the i-th and j-th AO coefficients on the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital, respectively, and $\varepsilon_{\mathrm{LUMO}}$ and $\varepsilon_{\mathrm{HOMO}}$ are the energies of these orbitals, which probably reflect involvement of the sulphonyl and carboxyl functions with the receptor site in the sweetness recognition.

TABLE VI

Results of correlations for α-arylsulfonylalkanoic acids (n = 10)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.762	25.6	0.178	0	7.93	0.64	12.2946	Intercept
				1	–1075	2127	–5.0579	Principal moment of inertia of C/# of atoms
2	0.941	55.3	0.096	0	5.01	0.11	45.5594	Intercept
				1	–30935	4123	–7.5037	Min. electroph. react. index for a C atom
				2	–2889	452	–6.3976	Min 1-electron react. index for a S atom

A previous literature study[10m] correlated the sweetness of these compounds employing a neural network and concluded that the *syn*-clinical conformation is probably actively involved in the display of the sweetness property.

Guanidines

The results of correlations for this sub-set of 27 guanidines (–N=C(NH)–NH–) are listed in Table VII. The three-parameter correlation (11) is more significant than the two-parameter correlation results according to the Fisher criteria. In Eq. (11), the moment of inertia I, the solvent-accessible surface area of H-bonding acceptor atoms, selected by threshold charge (HACA), and the minimum electron-electron repulsion for a C–C bond ($\mathrm{ER}_{\mathrm{min,CC}}$), defined as $\mathrm{ER}_{\mathrm{AB}} = \sum_{\mathrm{B}\neq\mathrm{A}} \sum_{\mu,\nu\in\mathrm{A}} \sum_{\lambda,\sigma\in\mathrm{B}} P_{\mu\nu} P_{\lambda\sigma} \langle\mu\nu|\lambda\sigma\rangle$, where A, B are atoms, $P_{\mu\nu}$ and $P_{\lambda\sigma}$ are the density matrix elements and $\langle\mu\nu|\lambda\sigma\rangle$ are the electron repulsion integrals on the atomic basis $\{\mu\nu\lambda\sigma\}$, are involved.

$$\begin{aligned}\ln \mathrm{RS} = &-(733 \pm 192)I - (0.181 \pm 0.03)\mathrm{HACA} - \\ &(0.682 \pm 0.148)\mathrm{ER}_{\mathrm{min,CC}} + (91.5 \pm 16.5) \qquad (11)\\ R^2 = 0.802 \quad &F = 31.1 \quad s = 0.215 \quad n = 27\end{aligned}$$

While the moment of inertia is related to the shape of the molecule, the other two factors probably influence the probability and enthalpy of the sweetener docking. The electron-electron repulsion energy describes the electron repulsion driven processes in the molecule and may be related to the conformational (rotational, inversional) changes or atomic reactivity in the molecule. No previous correlation has been located in the literature for the relative sweetness of guanidine sweeteners.

TABLE VII

Results of correlations for guanidines (n = 27)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.513	26.4	0.323	0	3.01	1.41	2.1282	Intercept
				1	1.28	0.25	5.1330	Kier & Hall index (order 2)
2	0.715	30.2	0.252	0	95.3	18.2	5.2511	Intercept
				1	–671	102	–6.6070	*RT*:1HACA2/TMSA
				2	–0.733	0.161	–4.5507	Min e-e repulsion for a C–C bond
3	0.802	31.1	0.215	0	91.5	16.5	5.5412	Intercept
				1	–733	192	–3.8155	Moment of inertia C
				2	–0.181	0.030	–6.0028	HACA-1 [Semi-MO PC]
				3	–0.682	0.148	–4.6014	Min e-e repulsion for a C–C bond

In several correlations reported in the present paper, the quantum-chemical energy descriptors have appeared as significant contributors. In each series of compounds, they are related to a specific bond or bond type, which could be expected to participate in the biological activity process. Proceeding from the structure of the Hartree-Fock Hamiltonian, the one- and two-atom energies can be further partitioned into coulombic and exchange (or resonance) terms. This corresponds to the fact that some processes (or interactions) are driven by the changes in the coulombic potential whereas others are directed by changes in the exchange potential. Alternatively, the processes may be controlled by a change in either the electronic or the nuclear energy[9]. Therefore, such partitioning of the quantum-chemical total energy for the series of molecules could potentially give additional valuable information about the physical character of processes involved in the substrate-receptor interactions. In Eq. (11), for instance, the significant contribution of the minimum electron-electron repulsion for a C-C bond ($\mathrm{ER}_{\mathrm{min,CC}}$) suggests the involvement of this bond during the interaction of guanidines with the sweetness receptor(s), presumably controlled by electronic repulsion energy.

Ureas and Thioureas

Correlations of the relative sweetness potency of a sub-set of sweeteners comprising 30 ureas (-NH-CO-NH-) and thioureas (-NH-CS-NH-) are listed in Table VIII. The best three-parameter correlation (Eq. (12)) is statistically significant (Table VIII) in view of the 30 data points available in this sub-set.

$$\ln \mathrm{RS} = (3.34 \pm 0.26)S_{\mathrm{trans}} - (5671 \pm 610)\mathrm{ZPC}_{\mathrm{max,H}} - (183 \pm 35)\mathrm{BO}_{\mathrm{avg,H}} + (585 \pm 75) \tag{12}$$

$$R^2 = 0.888 \quad F = 68.4 \quad s = 0.106 \quad n = 30$$

In Eq. (12), the descriptors involved are the translational entropy (S_{trans}), the maximum partial charge ($\mathrm{ZPC}_{\mathrm{max,H}}$) and the average bond order ($\mathrm{BO}_{\mathrm{avg,H}}$) for a hydrogen atom. The translational entropy is defined as,

$$S_{\mathrm{trans}} = \ln\left(\frac{2\pi mkT}{h^2}\right)^{1/2} \frac{Ve^{5/2}}{N} ,$$

where V is the volume of the system and N is the Avogadro's number. ZPC is the partial atom charge calculated using Zefirov's algorithm of the Sanderson electronegativity scheme.[8b] The Mulliken bond order[8b] for a given pair of atomic species A and B in the molecule is defined as

TABLE VIII

Results of correlations for ureas and for thioureas (n = 30)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.645	50.8	0.181	0	13.8	0.8	17.7640	Intercept
				1	–47.4	6.6	–7.1290	RPCG Relative positive charge (QMPOS/QTPLUS) [Semi-MO PC]
2	0.769	44.9	0.149	0	369	68	5.4081	Intercept
				1	0.310	0.034	9.0850	Structural Information Content (order 1)
				2	–3813	707	–5.3929	Max partial charge for a H atom [Zefirov's PC]
3	0.888	68.4	0.106	0	585	74.8	7.8241	Intercept
				1	3.34	0.26	12.8527	Translational entropy (300 K)
				2	–5671	610	–9.2928	Max partial charge for a H atom [Zefirov's PC]
				3	–183	35	–5.2511	Average bond order of a H atom
4	0.927	79.8	0.087	0	770	71	10.7783	Intercept
				1	–5644	510	–11.0739	Max partial charge (Qmax) [Zefirov's PC]
				2	$-4.41\text{x}10^{-2}$	$2.8\text{x}10^{-3}$	–15.6206	Total molecular 2-center exchange energy
				3	–238	32	–7.3536	Average bond order of a H atom
				4	–0.541	0.121	–4.4630	CSA-2 of C atoms

$$\mathrm{BO}_{\mathrm{AB}} = \sum_{i=1}^{\mathrm{occ}} \sum_{\mu\in \mathrm{A}} \sum_{\nu\in \mathrm{B}} n_i c_{i\mu} c_{j\nu} \quad ,$$

where the first summation is performed over all occupied molecular orbitals (n_i denotes the occupation number of the i-th MO), and the two other summations over μ and ν, the atomic orbitals belonging to the atoms A and B (A≠B) in the molecule. MO coefficients are denoted as $c_{i\mu}$ and $c_{j\nu}$.

The involvement of the entropy factor and the other charge related and bond order descriptors can be associated with the free energy pertinent to the docking. For this sub-set also, no previous literature correlation is available for comparison.

Natural Sweeteners

This sub-set comprises naturally occurring substances including sugars and their derivatives. For this sub-set of 20 compounds, correlation analysis (Table IX) afforded the following two-parameter Eq. (13).

$$\ln \mathrm{RS} = -(315 \pm 36)\mathrm{ER}_{\mathrm{max,O}} - (31.5 \pm 4.0)V_{XYZ,R} + (16.3 \pm 1.2) \quad (13)$$

$$R^2 = 0.905 \quad F = 81.1 \quad s = 0.219 \quad n = 20$$

In this equation, $\mathrm{ER}_{\mathrm{max,O}}$ and $V_{XYZ,R}$ correspond respectively to the maximum electrophilic reactivity index for an oxygen atom and the van der Waals molecular volume releated to volume of *XYZ* box for inertia moment oriented molecule.[8b] The descriptor $\mathrm{ER}_{\mathrm{max,O}}$, having inverse relationship with the nucleophilicity, has a negative sign for its coefficient, which may reflect the hydrogen bonding ability of the oxygen functionality. The second descriptor, $V_{XYZ,R}$ is related to the shape of the molecule, and also has a negative coefficient. Attempts to use the only volume related descriptors lead to poorer correlation with $R^2 = 0.4543$.

In a previous work, a relationship (14) between the apparent molar volume (AMV) and GEPOL/87 volume functions for carbohydrates was reported,[5a] but this study also does not correlate molecular descriptors directly with the sweetness potency.

$$\mathrm{AMV} = (0.750 \pm 0.025)V_{\mathrm{m}} - (1.83 \pm 4.4) \quad (14)$$

TABLE IX

Results of correlations for natural products ($n = 20$)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.782	64.4	0.327	0	5.85	0.38	15.2832	Intercept
				1	–2.06	0.26	–8.0250	LUMO+1 energy
2	0.905	81.1	0.219	0	16.3	1.2	13.0850	Intercept
				1	–315	36	–8.6383	Max electroph. react. index for an O atom
				2	–31.5	4.0	–7.9053	Molecular volume/ *XYZ* box
3	0.920	61.4	0.208	0	–4.47	2.49	–1.7919	Intercept
				1	10.7	1.2	9.0525	Average Information content (order 0)
				2	36.8	5.2	7.0985	Min net atomic charge
				3	0.347	0.11	3.1622	Kier flexibility index

Miscellaneous Sweeteners

For this group consisting of the remaining 7 sweeteners (Figure 1), which do not fit with any of the sub-sets discussed above, a one-parameter correlation (15) involving translational entropy is obtained (Table X).

Because of a lack of discriminating features for this subset, it is not possible to carry out a literature search for previous correlations.

$$\ln \mathrm{RS} = (2.59 \pm 0.25) S_{\mathrm{trans}} - (101 \pm 10) \qquad (15)$$

$$R^2 = 0.957 \quad F = 112 \quad s = 0.266 \quad n = 7$$

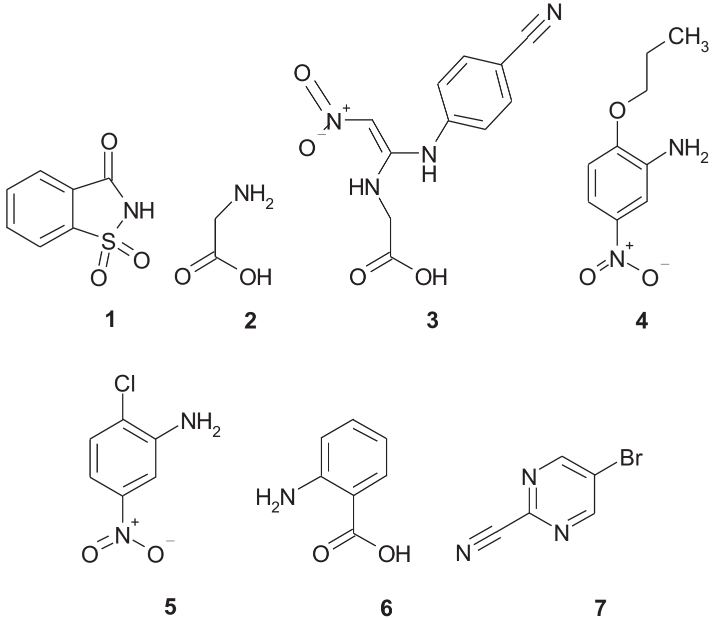

Figure 1. Structures of miscellaneous sweeteners.

Peptides

The sweetness of peptides depends on their stereochemical structure. As only (active) L,L-stereoisomers were included the set considered, the resulting QSPR equations are valid only for this stereochemical class of compounds. The sweetness potency of this sub-set of 87 sweeteners with peptide bonds (–NH–CO–) was analysed by the heuristic and the best multi-linear regression methods. The best correlations (Table XI) include the statistically significant four- and five-parameter Eqs. (16) and (17).

TABLE X

Results of correlations for unclassified structures (n = 7)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.957	112	0.264	0	–101	10	–10.03	Intercept
				1	2.59	0.24	10.60	Translational entropy
2	0.997	647	0.079	0	14.5	0.6	–22.6451	Intercept
				1	4.83	0.2	26.0874	Average Information content (order 2)
				2	145	9	15.4931	Maximum 1-electron reaction index for a C atom

$$\ln \mathrm{RS} = -(6.52 \pm 0.61) \times 10^{-4} H_{\mathrm{tot}} + (43.7 \pm 10.0)\mathrm{ZPC}_{\mathrm{max}} + (1.54 \pm 0.47)\#\mathrm{Cl} + (16.8 \pm 5.5) q_{\mathrm{min,N}} - (2.44 \pm 2.59) \qquad (16)$$

$$R^2 = 0.649 \quad F = 37.8 \quad s = 0.145 \quad n = 87$$

$$\ln \mathrm{RS} = (7.48 \pm 0.71) \times 10^{-2} S_{\mathrm{tot}} - (47.4 \pm 9.5)\mathrm{ZPC}_{\mathrm{max}} + (1.87 \pm 0.45)\#\mathrm{Cl} + (0.630 \pm 0.184) E_{\mathrm{en,CO}} + (18.0 \pm 5.5) q_{\mathrm{min,N}} - (266 \pm 76) \qquad (17)$$

$$R^2 = 0.688 \quad F = 35.7 \quad s = 0.138 \quad n = 87$$

The descriptors in Eqs. (16) and (17) are the total enthalpy[8b] (H_{tot}), Zefirov's maximum partial charge ($\mathrm{ZPC}_{\mathrm{max}}$), the number of chlorine atoms (#Cl), Mulliken's minimum net atomic charge[8b] for a nitrogen atom ($q_{\mathrm{min,N}}$), the total entropy (S_{tot}), and the maximum electron-neutron attraction for a carbon-oxygen bond ($E_{\mathrm{en,CO}}$). The enthalpy and entropy factors presumably model the free energy of docking the sweetener with the receptor site. The other descriptors may reflect hydrogen bonding and electrostatic interactions involved in docking.

The plurality of conformations and the lack of descriptors for taking into account the absolute configuration of the stereocenters in the many chiral sweeteners in this sub-set could preclude better correlations.

The only[5a] previously reported correlation for peptide sweeteners was the use of GEPOL/87 computer program to predict apparent specific volume (ASV) and apparent molar volume (AMV) which, in turn, allow quantitative prediction of sweetness. The sweet taste corresponds to ASV values in the range of 0.52–0.71,[5a] and a good correlation was found between AMV and the molecular volume: AMV = $(0.811 \pm 0.045)V_{\mathrm{m}} - (11.4 \pm 5.8)$ ($r = 0.978$) for 17 amino acids. However, this correlation does not constitute a structure-taste relationship.

TABLE XI

Results of correlations for peptides (n = 87)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.483	79.5	0.172	0	–6.66	1.48	–4.4965	Intercept
				1	0.749	0.084	8.9167	Information content (order 0)
2	0.573	56.4	0.155	0	–13.3	1.9	–6.8380	Intercept
				1	$8.04\mathrm{x}10^{-2}$	$7.9\mathrm{x}10^{-3}$	10.1568	Total entropy
				2	52.7	10.7	4.9137	Maximum partial charge (Qmax) [Zefirov's PC)
3	0.619	45.0	0.152	0	–13.0	1.8	–7.0514	Intercept
				1	$-8.22\mathrm{x}10^{-2}$	$7.5\mathrm{x}10^{-3}$	10.8955	Total entropy
				2	46.4	10.4	4.4720	Maximum partial charge (Qmax) [Zefirov's PC)
				3	1.52	0.48	3.1640	Number of Cl atoms
4	0.649	37.8	0.145	0	–2.44	2.59	–0.9412	Intercept
				1	$-6.52\mathrm{x}10^{-4}$	$6.1\mathrm{x}10^{-5}$	10.7537	Total enthalpy
				2	43.7	10.0	4.3633	Maximum partial charge (Qmax) [Zefirov's PC)
				3	1.54	0.47	3.3093	Number of Cl atoms
				4	16.8	5.5	3.0613	Minimum net atomic charge for a N atom
5	0.689	35.7	0.138	0	–266	76	–3.5239	Intercept
				1	$7.48\mathrm{x}10^{-2}$	$7.1\mathrm{x}10^{-3}$	10.4867	Total entropy
				2	47.4	9.5	4.9719	Maximum partial charge (Qmax) [Zefirov's PC)
				3	1.87	0.45	4.1696	Number of chlorine atoms
				4	0.630	0.184	3.4243	Maximum e-n attraction for a C-O bond
				5	18.03	5.46	3.3030	Minimum net atomic charge for a N atom

Later, the same group reviewed[5b] the molar volume contribution to the sweet taste and proposed an association of sweet taste with another solution property *viz*. the partial molar isentropic compressibility which defines the compactness of the hydration layer around sweet molecules. This study did not propose any structure-sweetness correlation.

Global Set of Sweeteners with Quantitative Sweetness Potencies

Before discussing global correlations covering the whole set of sweeteners, the question of whether there exists single or multiple sweetness receptors must be considered. This is contraversial. Nofre *et al.*[7a] suggest a single receptor with multiple recognition sites and that different classes of sweetness contain different numbers of binding sites. However, Linderman and co-workers[7b] have demonstrated rather convincingly that different sweeteners have different signaling pathways, and Shaeffer *et al.*[7c] provide sensory evidence for multiple receptors. Sugars generally probably have a different mode of action than high potency sweeteners.[6b] Further complications include the fact that only some sweeteners have synergy, the taste can be observed at different part of the tongue-front or middle, and one can block the sweetness of some compounds. Nevertheless, even if different classes of compounds interact with different receptors, their biological and structural similarity of them may lead to similar interactions governed by similar descriptors. The results obtained in Eqs. (18) – (21) with significant correlation coefficients seem to indicate that similar interactions are responsible for the sweetness of structurally different compounds.

Correlation analysis of the general set of 238 sweeteners using linear regression methods (Table XII) provides significant four- and five-parameter correlation Eqs. (18) and (19).

$$\begin{gathered}\ln \mathrm{RS} = (1.59 \pm 0.12) \times 10^{-3}\mathrm{GI} + (2.08 \pm 0.15) \times 10^{-2} F_{\mathrm{T}} - \\ (0.177 \pm 0.034)\#\mathrm{B}_{\mathrm{aromatic}} + (18.0 \pm 4.5)\mathrm{ZPC}_{\max,\mathrm{C}} + (1.05 \pm 0.31) \qquad (18) \\ R^2 = 0.671 \quad F = 119 \quad s = 0.108 \quad n = 238\end{gathered}$$

$$\begin{gathered}\ln \mathrm{RS} = (1.38 \pm 0.13) \times 10^{-3}\mathrm{GI} + (1.57 \pm 0.21) \times 10^{-2} F_{\mathrm{T}} - \\ (0.144 \pm 0.035)\#\mathrm{B}_{\mathrm{aromatic}} + (22.6 \pm 4.6)\mathrm{ZPC}_{\max,\mathrm{C}} - \\ (34.9 \pm 10.7)\mathrm{PSA}_{\mathrm{O}} + (2.14 \pm 0.45) \qquad (19) \\ R^2 = 0.686 \quad F = 101 \quad s = 0.098 \quad n = 238\ .\end{gathered}$$

Eqs. (18) and (19) involve GI (gravitation index, all pairs[8b]), F_{T} (thermodynamic heat of formation of the molecule), $\#\mathrm{B}_{\mathrm{aromatic}}$ (number of aromatic bonds), $\mathrm{ZPC}_{\max,\mathrm{C}}$ (maximum partial charge for a carbon atom) and $\mathrm{PSA}_{\mathrm{O}}$ (positively charged solvent accessible surface area[8b] of oxygen atoms). The gravitational index defined as

TABLE XII

Results of correlations for the general set of structures (n = 238)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.443	188	0.140	0	0.567	0.421	1.3460	Intercept
				1	3.65×10^{-2}	2.7×10^{-3}	13.6998	α polarizability
2	0.616	188	0.116	0	0.440	0.364	1.2083	Intercept
				1	2.52×10^{-2}	1.3×10^{-3}	18.7386	Molecular weight
				2	0.0153	0.0013	12.1751	Thermodynamic heat of formation of the molecule
3	0.648	144	0.112	0	1.492	0.297	5.0172	Intercept
				1	1.77×10^{-3}	1.1×10^{-4}	16.0325	Gravitation Index (all pairs)
				2	2.08×10^{-2}	1.5×10^{-3}	13.8246	Thermodynamic heat of formation of the molecule
				3	–0.168	0.035	–4.7360	Number of aromatic bonds
4	0.671	119	0.108	0	1.05	0.308	3.4198	Intercept
				1	1.59×10^{-3}	1.2×10^{-4}	13.6032	Gravitation Index (all pairs)
				2	2.08×10^{-2}	1.5×10^{-3}	14.2434	Thermodynamic heat of formation of the molecule
				3	–0.177	3.4×10^{-2}	–5.1347	Number of aromatic bonds
				4	18.0	4.5	4.0078	Max partial charge for a carbon atom [Zefirov's PC]
5	0.686	101	0.105	0	2.14	0.45	4.7711	Intercept
				1	1.38×10^{-3}	1.3×10^{-4}	10.5752	Gravitation Index (all pairs)
				2	1.57×10^{-2}	2.1×10^{-3}	7.4524	Thermodynamic heat of formation of the molecule
				3	–0.144	3.5×10^{-2}	–4.0858	Number of aromatic bonds
				4	22.6	4.6	4.9000	Maximum partial charge for a C atom [Zefirov's PC]
				5	–34.9	10.7	–3.2741	PSA-2 of Oxygen atoms

$$\mathrm{GI_p} = \sum_{A<B}^{N_p} \frac{m_A m_B}{r_{AB}^2} ,$$

where m_A and m_B are the atomic masses of atoms A and B, r_{AB} is interatomic distance, and N_p is number of atoms in molecule. The involvement of all these descriptors has understandable physicochemical meaning as described below.

The first descriptor, *viz*. the gravitation index, can be linked to the extent of docking. The descriptor coefficient for the number of aromatic bonds has negative sign which probably indicates that the rigidity of the aromatic rings renders the docking of the sweetener difficult. The maximum partial charge for the carbon atom has a positive descriptor coefficient indicating that an increase in the charge facilitates the interaction with the receptor protein.

The positively charged solvent accessible surface area of oxygen goes against the ability of oxygen to form hydrogen bonds with the protein molecules and hence an increase in the value of this descriptor leads to decreased sweetness as evident from the negative sign of the descriptor coefficient. Figure 2 depicts the comparison of the observed ln RS *versus* predicted ln RS values for this global set on the basis of five-parameter equation. There is no literature report available on the structure-sweetness correlation for the global set of sweeteners.

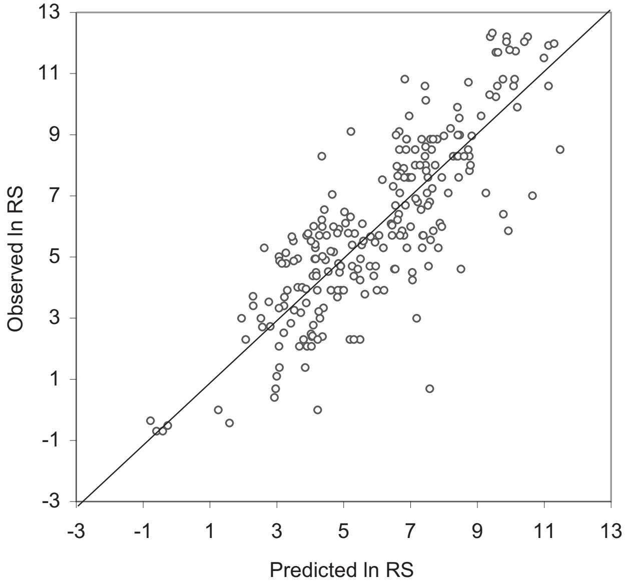

Figure 2. Plot of observed ln RS *versus* predicted ln RS for the whole set (n = 238) using five-parameter equation (19) (R^2 = 0.686).

Global Set of Sweeteners Excluding the Peptide Sub-set

It was considered worthwhile to investigate whether the correlation for the general set can be improved by deleting one or more sub-sets. Such an improvement in correlation, in turn, can furnish additional, more reliable information on the nature of factors involved in determining the sweetness potency of the general set, besides the possibility of bringing out the sub-set's unique features pertinent to deviations to light. Many of the outliers (63 structures, outside $\pm 2\sigma$, 95% reliability level) belong to three groups: the aldoximes in the area of low sweetness, peptides in the medium area, and the guanidines in the area of high sweetness (Table XIII). This is indirect evidence for the existence of multiple receptor sites. A correlation was performed on the general set excluding the largest sub-set, *viz.* 87 peptides sweeteners and the results are listed in Table XIV. The results show that all the correlations (both single and multi-parameter) are significantly improved for the general set lacking peptides.

The data in Table XIV are obtained from the following four- and five-parameter correlations (20 and 21 respectively), wherein structural information content (order 0), thermodynamic heat of formation of the molecule, polarity parameter ($Z_{\max} - Z_{\min}$) and HOMO-1 energy are involved.

TABLE XIII

Outliers (Eq. 19), sorted by ascending of experimental ln RS

#	Class	Exp. ln RS	Calc. ln RS	Diff.
S0240	Miscellaneous	1.585	–0.431	2.016
S0184	Aldoxime	2.623	5.298	–2.675
S0060	Aldoxime	2.927	0.406	2.521
S0061	Aldoxime	2.965	0.693	2.272
S0079	Aldoxime	2.996	1.099	1.898
S0187	Aldoxime	3.064	5.011	–1.947
S0276	Peptide	3.270	5.136	–1.865
S0205	Natural	3.451	5.666	–2.215
S0186	Aldoxime	3.494	5.521	–2.027
S0057	Aldoxime	3.848	1.386	2.462
S0183	Miscellaneous	3.892	5.704	–1.811
S0076	Aldoxime	3.904	2.079	1.825
S0275	Peptide	3.934	5.768	–1.834
S0095	Aldoxime	4.036	2.079	1.957
S0084	Aldoxime	4.104	6.009	–1.904
S0390	Peptide	4.220	0.000	4.220
S0315	Urea	4.225	2.303	1.923
S0054	Aldoxime	4.346	6.215	–1.869

TABLE XIII (Cont.)

S0299	Miscellaneous	4.347	8.294	–3.947
S0100	α–Sulfonyl acid	4.365	2.398	1.967
S0304	Urea	4.423	6.551	–2.127
S0048	Aldoxime	4.660	7.048	–2.387
S0075	Aldoxime	5.184	2.303	2.882
S0212	Natural	5.219	9.105	–3.885
S0393	Peptide	5.305	2.303	3.003
S0089	Aldoxime	5.493	2.303	3.191
S0104	α–Sulfonyl acid	5.635	3.784	1.851
S0269	Peptide	5.997	3.912	2.086
S0214	Natural	6.208	3.912	2.297
S0286	Peptide	6.521	4.605	1.916
S0409	Peptide	6.545	4.605	1.940
S0268	Peptide	6.569	8.987	–2.418
S0249	Peptide	6.659	9.105	–2.446
S0415	Peptide	6.669	8.517	–1.848
S0337	Peptide	6.832	10.820	–3.988
S0416	Peptide	6.881	8.854	–1.973
S0253	Peptide and urea	6.884	8.854	–1.969
S0250	Urea	6.961	9.616	–2.655
S0399	Peptide	7.056	4.500	2.556
S0402	Peptide	7.059	4.248	2.811
S0422	Peptide	7.181	2.996	4.185
S0410	Peptide	7.350	5.298	2.052
S0342	Peptide	7.434	10.597	–3.162
S0357	Urea	7.458	10.127	–2.669
S0401	Peptide	7.541	4.700	2.841
S0403	Peptide	7.573	0.693	6.880
S0400	Peptide	7.600	5.560	2.039
S0244	Guanidine	7.683	5.858	1.825
S0411	Peptide	7.831	5.298	2.533
S0424	Peptide	7.938	5.991	1.946
S0363	Guanidine	8.509	4.605	3.904
S0317	Guanidine	8.735	10.714	–1.989
S0044	Peptide	9.257	7.090	2.167
S0260	Guanidine	9.383	12.206	–2.823
S0262	Guanidine	9.449	12.324	–2.874
S0258	Guanidine	9.554	11.695	–2.141
S0333	Guanidine	9.614	11.695	–2.081
S0226	Miscellaneous	9.777	6.397	3.380
S0344	Guanidine	9.868	12.044	–2.176
S0334	Guanidine	9.878	12.206	–2.328
S0364	Guanidine	9.925	5.858	4.067
S0321	Guanidine	9.964	11.775	–1.812
S0227	Guanidine	10.649	7.003	3.646
S0324	Guanidine	11.477	8.517	2.960

TABLE XIV

Results of correlations for general set excluding peptide sub-set (n = 151)

#P	R^2	F	s		B	ΔB	t	Name of descriptor
1	0.599	223	0.169	0	1.05	0.37	2.8208	Intercept
				1	0.176	0.012	14.9261	SA-2 of C atoms
2	0.770	248	2.487	0	0.147	0.341	0.4317	Intercept
				1	$-1.54\text{x}10^{-3}$	$7\text{x}10^{-5}$	–21.0229	Total molecular 1-center E-N attraction
				2	$2.34\text{x}10^{-2}$	$1.4\text{x}10^{-3}$	17.1355	Thermodynamic heat of formation of the molecule
3	0.794	189	0.128	0	5.98	1.01	5.9399	Intercept
				1	0.371	0.017	22.1669	Structural information content (order 1)
				2	$1.31\text{x}10^{-2}$	$1.2\text{x}10^{-3}$	11.2776	Thermodynamic heat of formation of the molecule
				3	–2.89	4.04	–7.1596	Polarity parameter ($Q_{\max}$-$Q_{\min}$) (Zefirov's PC)
4	0.820	166	0.114	0	18.8	3.2	5.9551	Intercept
				1	0.507	0.038	13.3883	Structural information content (order 0)
				2	$9.89\text{x}10^{-3}$	$1.18\text{x}10^{-3}$	8.3961	Thermodynamic heat of formation of the molecule
				3	–25.9	4.05	–6.4003	Polarity parameter ($Q_{\max}$-$Q_{\min}$) [Zefirov's PC]
				4	1.22	0.23	5.3537	HOMO-1 energy
5	0.842	154	0.108	0	18.3	2.98	6.1389	Intercept
				1	0.50	0.036	14.0190	Structural information content (order 0)
				2	$1.03\text{x}10^{-2}$	$1.1\text{x}10^{-3}$	9.2400	Thermodynamic heat of formation of the molecule
				3	1.23	0.27	4.4696	Number of chlorine atoms
				4	–23.7	3.8	–6.1694	Polarity parameter ($Q_{\max}$-$Q_{\min}$) [Zefirov's PC]
				5	1.22	0.21	5.6832	HOMO-1 energy

$$\ln \mathrm{RS} = (0.507 \pm 0.038)\,{}^{0}\mathrm{SIC} + (9.89 \pm 0.118) \times 10^{-2} F_{\mathrm{T}} - (25.9 \pm 4.0)(\mathrm{ZPC}_{\max} - \mathrm{ZPC}_{\min}) + (1.22 \pm 0.23)\varepsilon_{\text{HOMO-1}} + (18.8 \pm 3.2) \quad (20)$$

$$R^2 = 0.820 \quad F = 166 \quad s = 0.114 \quad n = 151$$

$$\ln \mathrm{RS} = (0.500 \pm 0.036)\,{}^{0}\mathrm{SIC} + (1.03 \pm 0.11) \times 10^{-2} F_{\mathrm{T}} - (1.23 \pm 0.27)\#\mathrm{Cl} - (23.7 \pm 3.8)(\mathrm{ZPC}_{\max} - \mathrm{ZPC}_{\min}) + (1.22 \pm 0.21)\varepsilon_{\text{HOMO-1}} + (18.3 \pm 3.0) \quad (21)$$

$$R^2 = 0.842 \quad F = 154 \quad s = 0.108 \quad n = 151$$

The structural information content[8b] (^{0}SIC) is a topological index is based on Shannon information theory and is defined as:

$$ {}^{k}\mathrm{SIC} = \frac{{}^{k}\mathrm{IC}}{\log_2 n},$$

$$ {}^{k}\mathrm{IC} = -\sum_{i=1}^{k} \frac{n_i}{n} \log_2 \frac{n_i}{n}$$

where n_i is a number of atoms in the i-th class, n is the total number of atoms in the molecule, and k is the number of atomic layers in the coordination sphere around a given atom that is accounted for (in our case $k = 0$).

The significance of these parameters can be explained in line with the influence of several descriptors, discussed in the previous sets, on the free energy of the docking of the sweetener with the protein receptors.

Nature of Interactions Involved in the Sweetening Process

The descriptors emerged from the foregoing correlations are classified and tabulated in Table XV with a view to identifying the important factors contributing to sweetness. Quantum chemical descriptors appear most frequently (15 times), followed by electrostatic (11 times), constitutional (9 times), thermodynamical (8 times), geometrical (3 times) and topological (2 times). The influence of these descriptors on the sweetness potency varies as is evident from the relative magnitudes of the descriptor coefficients in the correlation equations. The constitutional, topological, geometrical, surface- and/or size- related descriptors may influence the binding of the sweetener with the receptor, while the electronic and charge related descriptors may influence the chemical reactivity and/or the electrostatic interactions between the sweetener and the receptor.

TABLE XV

Descriptors encountered in the correlation equations of sweetness potency

S/No	Name of descriptor	Occurring equation
	Constitutional	
1	Number of oxygen atoms	1, 2
2	Number of chlorine atoms	16, 17, 21
3	Gravitation index	18, 19
4	Number of aromatic bonds	18, 19
	Topological	
5	Structural information content (order 0)	20, 21
	Geometrical	
6	Solvent accessible surface area of carbon	1
7	Moment of Inertia C	11
8	Molar volume / *XYZ* Box	13
	Electrostatic	
9	Positively charged solvent accessible surface area of carbon	2
10	Hydrogen acceptor solvent accessible charged surface area	11
11	Maximum partial charge for a hydrogen atom	12, 17
12	Maximum partial charge	16, 17
13	Maximum partial charge for a carbon atom	18, 19
14	Positively charged solvent accessible surface area of oxygen	19
15	Polarity parameter	20, 21
	Quantum chemical	
16	Positively charged solvent accessible surface area of carbon	2
17	Average electrophilic reaction index of carbon	2
18	Maximum coloumbic interaction for the carbon-hydrogen bond	6
19	Maximum net atomic charge for a hydrogen atom	7
20	Minimum resonance energy for a carbon-chlorine bond	7
21	Minimum electrophilic reaction index for a carbon atom	10
22	Minimum one electron reaction index for a sulphur atom	10
23	Minimum electron-electron repulsion for a carbon-carbon bond	11
24	Average bond order for a hydrogen atom	12
25	Maximum electrophilic reaction index for oxygen	13
26	Minimum net atomic charge for a nitrogen atom	16, 17
27	Maximum electron-neutron attraction for a carbon-oxygen bond	17
28	Energy of the highest occupied molecular orbital*	20, 21
	Thermodynamic	
29	Translational entropy	12, 15
30	Total enthalpy	16
31	Total entropy	17
32	Thermodynamic heat of formation of the molecule	18,19,20,21

CONCLUSIONS

The development of significant QSAR or QSPR equations by extraction of molecular descriptors from large descriptor spaces has assisted in the prediction of many physical properties and biological activities of chemical compounds. The present work demonstrates that analogous QSPR equations can be developed for the description of the complex biochemical processes of sweetness. The descriptors employed in the best correlation equations can indicate the interactions involved in the sweet taste mechanism and the underlying docking process.

It is possible that a unique blend of factors could influence the sweetness property of different classes of sweeteners because of the variation in the nature of the functionality and the extent of interaction with the receptor in docking, despite the possible involvement of different receptors with different functionalities. Interestingly, descriptors pertinent to enthalpy, entropy, polarity, shape, charge distribution and structure are those often encountered in different types of sweeteners as factors that determine the sweetness potency.

Acknowledgement. – We thank the NutraSweet Company for support of this work.

REFERENCES

1. R. M. Horowitz and B. Gentili, *Dihydrochalcone Sweeteners from Citrus Flavanones,* in: L. O'. Nabors and R. C. Gelardi (Eds.), *Alternative Sweeteners,* Marcel and Dekker Inc., New York, 1986, pp. 135–153.
2. (a) Y. Takahashi, Y. Miyashita, Y. Tanaka, H. Hayasaka, H. Abe, and S.-I. Sasaki, *J. Pharm. Sci.* **73** (1984) 737–741. (b) H. Iwamura, *J. Med. Chem.* **23** (1980) 308–312.
3. (a) Y. Miyashita, Y. Takahashi, C. Takayama, T. Ohkubo, K. Funatsu, and S.-I. Sasaki, *Anal. Chim. Acta* **184** (1986) 143–149. (b) W. J. Spillane and M. B. Sheahan, *J. Chem. Soc., Perkin Trans. 2* (1989) 741–746.
4. (a) S. Shamil, G. G. Birch, M. Dinovi, and R. Rafka, *Food Chemistry* **32** (1989) 171–188. (b) W. J. Spillane, G. Morini, and G. G. Birch, *Food Chemistry* **44** (1992) 337–342.
5. (a) W. J. Spillane, G. G. Birch, M. G. B. Drew, and I. Bartolo, *J. Chem. Soc., Perkin Trans. 2* (1992) 497–503. (b) G. G. Birch, S. Parke, R. Siertsema, and J. M. Westwell, *Pure & Appl. Chem.* **69** (1997) 685–692.
6. (a) S. S. Schiffman, *Pure & Appl. Chem.* **69** (1997) 701–708. (b) D. J. Ager, D. P. Pantaleone, S. A. Henderson, A. R. Katritzky, I. Prakash, and D. E. Walters, *Angew. Chem.* **37** (1998) 1802–1817. (c) A. van der Heijden, *Pure & Appl. Chem.* **69** (1997) 667–674.
7. (a) C. Nofre and J.-M. Tinti, *Food Chemistry* **56** (1996) 263–274. (b) S. J. Bernhardt, M. Naim, U. Zehavi, and B. Lindermann, *J. Physiol*. **490.2** (1996) 325–336.

(c) S. S. Schiffman, H. Cahn, and M. G. Lindley, *Pharmacol. Biochem. & Behavior* **15** (1981) 377–388.

8. (a) A. R. Katritzky, V. S. Lobanov, and M. Karelson, *CODESSA Reference Manual,* Version 2.0, University of Florida, 1996. (b) M. Karelson, *Molecular Descriptors in QSAR/QSPR,* Wiley-Interscience, New York, 2000. (c) A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, *J. Chem. Inf. Comp. Sci.* **40** (2000) 1–18. (d) M. Karelson, U. Maran, Y. L. Wang, and A. R. Katritzky, *Collect. Czech. Chem.* **64** (1999) 1551–1571. (e) A. R. Katritzky, L. Mu, V. S. Lobanov, and M. Karelson, *J. Phys. Chem.* **100** (1996) 10400–10407. (f) A. R. Katritzky, V. S. Lobanov, and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38** (1998) 28–41. (g) A. R. Katritzky, Y. L. Wang, S. Sild, T. Tamm, and M. Karelson, *J. Chem. Inf. Comp. Sci.* **38** (1998) 720–725. (h) A. R. Katritzky, T. Tamm, Y. Wang, S. Sild, and M. Karelson, *J. Chem. Inf. Comput. Sci.* **39** (1999) 684–691. (i) A. R. Katritzky, T. Tamm, Y. Wang, and M. Karelson, *J. Chem. Inf. Comp. Sci.* **39** (1999) 692–698. (j) A. R. Katritzky, S. Sild, and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38** (1998) 840–844. (k) A. R. Katritzky, L. Mu, and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38** (1998) 293–299. (l) P. D. T. Huibers, D. O. Shah, and A. R. Katritzky, *J. Colloid Interface Sci.* **193** (1997) 132–136. (m) A. R. Katritzky, E. S. Ignatchenko, R. A. Barcock, V. S. Lobanov, and M. Karelson, *Anal. Chem.* **66** (1994) 1799– 1807.
9. I. G. Csizmadia, *Theory and Practice of MO Calculations on Organic Molecules*, Elsevier, Amsterdam, 1976, p. 249.
10. (a) L. B. Kier, *J. Pharm. Sci.* **69** (1980) 416–418. (b) L. O. Nabors and G. E. Inglett, *A Review of Various Other Alternative Sweeteners,* in: L. O. 'Nabors and R. C. Gelardi (Eds.), *Alternative Sweeteners,* Marcel Dekker Inc., New York, 1986, pp. 309–323. (c) A. Ripper, B. E. Homler, and G. A. Miller, *Aspartame,* in: L. O.' Nabors and R. C. Gelardi (Eds.), *Alternative Sweeteners,* Marcel Dekker Inc., New York, 1986, pp. 43–70. (d) G.-W. von Rymon Lipinski, *Acesulfame-K,* in: L. O.' Nabors and R. C. Gelardi (Eds.), *Alternative Sweeteners,* Marcel Dekker Inc., New York, 1986, pp 89–102. (e) T. J. Venanzi and C. A. Venanzi, *Electrostatic Recognition Patterns of Sweet-Tasting Compounds,* in: E. D. Walters, F. T. Orthoefer, and G. E. Dubois (Eds), *Sweetener,* ACS Symposium Series 450, American Chemical Society, Washington, 1991, pp. 193–205. (f) R. C. Glowaky, M. E. Hendrick, R. E. Smiles, and A. Torres, *Development and Uses of Alitame. A Novel Dipeptide Amide Sweeteners,* in: E. D. Walters, F. T. Orthoefer, and G. E. Dubois, (Eds.), *Sweeteners,* ACS Symposium Series 450, American Chemical Society, Washington, 1991, pp. 57–67. (g) G. W. Muller, D. L. Madigan, J. Chris Culberson, D. E. Walters, J. S. Carter, C. A. Klade, G. E. DuBois, and M. S. Kellogg, *High Potency Sweeteners Derived from β-Amino Acids,* in: E. D. Walters, F. T. Orthoefer and G. E. Dubois (Eds.), *Sweeteners,* ACS Symposium Series 450, American Chemical Society, Washington, 1991, pp. 113–125. (h) D. C. Culberson and D. E. Walters, *Three-Dimensional Model for the Sweet Taste Receptor – Development and Use,* in: E. D. Walters, F. T. Orthoefer, and G. E. Dubois (Eds.), *Sweeteners,* ACS Symposium Series 450, American Chemical Society, Washington, 1991, pp. 214–223. (i) J.-M. Tinti and C. Nofre, *Design of Sweeteners. A Rational Approach,* in: E. D. Walters, F. T. Orthoefer, and G. E. Dubois (Eds.), *Sweeteners,* ACS Symposium Series 450, American Chemical Society, Washington, 1991, pp. 88–99. (j) J. W. Ellis, S. H. Malehorn, L. M. Browing, and T. A. Heischmidt, *J. Carbohydrate Chem.* **11** (1992) 761–778. (k) A. Arnoldi, A. Bassoli, and L. Merlini, *Food Chemistry* **56** (1996) 247–253. (l) D. Glaser, J.-M. Tinti, and C. Nofre, *Food Chemistry* **56** (1996)

313–321. (m) J. Polanski, A. Ratajczak, J. Gasteiger, Z. Galdecki, and E. Galdecka, *J. Mol. Struct.* **407** (1997) 71–80. (n) R.-H. Mattern, Y. Amino, E. Benedetti, and M. Goodman, *Biopolymers* **49** (1999) 525–539. (o) *US-Patent 5997933 Sweetening agent derived from* L*-aspartic or* L*-glutamic acid.* (p) *US-Patent-5374733* N*-(3,3-dimethylbutyl)-*L*-aspartyl-*D*-α-aminoalkanoic acid* N*-(*S*)-1-phenyl-1-alkanamide useful as a sweetening agent.*

11. *Integrated Scientific Information System* (documentation package), MDL Information Systems, Inc., 1997.
12. QCPE455. MOPAC: A General Molecular Orbital Package (Version 6.0)
13. E. V. Gordeeva, A. R. Katritzky, V. V. Shcherbukhin, and N. S. Zefirov, *J. Chem. Inf. Comp. Sci.* **33** (1993) 102–111.
14. http://www.hyper.com.
15. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* **107** (1985) 3902–3909.

SAŽETAK

Kvantitativna analiza odnosa strukture i stupnja slatkoće s pomoću programa CODESSA

Alan R. Katritzky, Ruslan Petrukhin, Subbu Perumal, Mati Karelson,
Indra Prakash i Nitin Desai

397 prirodnih i umjetnih sladila klasificirano je na temelju njihove strukture u devet skupina. Stupanj slatkoće koreliran je s kvantnokemijskim i drugim molekulskim deskriptorima upotrebom heurističkih i najboljih multilinearnih regresijskih postupaka iz programskog paketa CODESSA. Kvantitativni modeli odnosa strukture i svojstava (QSPR modeli) [ako nije drugačije označeno modeli sadržavaju dva deskriptora], dobiveni za svaku pojedinu klasu sladila, imaju sljedeće vrijednosti R^2: 0,835 za 47 aldoksima, 0,959 za 8 acesulfamata, 0,919 za 9 sulfamata, 0,941 za 10 α-arilsulfonilalkanoičnih kiselina, 0,715 za 27 guanidina (0,802 u korelaciji s tri parametra), 0,769 za 30 uree/tiouree (0,888 u korelaciji s tri parametara), 0,905 za 20 prirodnih sladila, 0,957 za 7 raznovrsnih sladila (korelacija s jednim parametrom), 0,688 za 87 peptida (korelacija s pet parametara). Dan je također i model za cjelokupni skup sladila, koji uključuje pet deskriptora (R^2 = 0,686), i koji je razmotren s obzirom na mogućnost postojanja jednoga ili više receptora za sladila.