

Rating of Sweetness by Molar Refractivity and Ionization Potential: QSAR Study of Sucrose and Guanidine Derivatives

Rajesh K. Singh*, Mohd. A. Khan and Pashupati P. Singh

Department of Chemistry, M.L.K. P.G. College, Balrampur, U.P., India.

Received 31 December 2012, revised 29 August 2013, accepted 19 December 2013.

ABSTRACT

A quantitative structure activity relationship study of 31 sucrose derivatives and 30 guanidine derivatives has been undertaken. Their sweetness values, relative to sucrose (RS), have been taken from literature. The study has been made with the help of CAChe Pro software by using eight descriptors, *viz.* electron affinity, ionization potential, electrophilicity index, total energy, heat of formation, steric energy, molar refractivity and solvent accessible surface area. Multi-linear regression (MLR) analysis has been performed with different combinations of descriptors and the quality of regression has been adjudged by the correlation coefficient, cross-validation coefficient and other statistical parameters like the standard error, standard error of the estimate, degrees of freedom, etc. The study indicates that ionization potential appears an important descriptor for sucrose derivatives, whereas molar refractivity appears an important descriptor for guanidine derivatives. The ionization potential alone and in combination with the electrophilicity index, molar refractivity and solvent accessibility surface area provide dependable QSAR models for sucrose derivatives. Molar refractivity alone and in combination with solvent accessibility surface area, ionization potential and heat of formation provide dependable QSAR models for guanidine derivatives. The predicted sweetness values obtained by these QSAR models are close to observed sweetness.

KEYWORDS

Sweetness, sucrose, guanidine, DFT, PM3.

1. Introduction

Sweetness is one of the basic tastes and is almost universally regarded as a pleasurable experience. Sucrose (table sugar) is the prototypical example of a sweet substance. Sucrose has a sweetness perception rating of 1 and other substances are rated relative to this. Sweet taste is thought to arise from the interaction of a molecule with a G-protein coupled taste receptor, identified as the T1R3 receptor at the taste receptor cells, which generate a sensation of pleasant sweetness.^{1–4} The study of the structure-sweet taste relationship was first systematically started with the development of the AH-B theory of sweetness proposed by Shallenberger and Acree.⁵ They proposed that to be sweet, a compound must contain a hydrogen bond donor (AH) and a Lewis base (B); the AH-B unit of a sweetener binds with a corresponding AH-B unit on the biological sweetness receptor to produce the sensation of sweetness. Later on the B-X theory was proposed by Lemont Kier.⁶ He proposed that to be sweet, a compound must have a third binding site (labelled X) that could interact with a hydrophobic site on the sweetness receptor *via* London dispersive forces. The most elaborate theory of sweetness to date is the multipoint attachment theory (MPA) proposed by Tinti and Nofre.^{7,8} This theory involves a total of eight interaction sites between a sweetener and the sweetness receptor, although not all sweeteners interact with all eight sites.

The main thrust of QSAR studies has been in the field of drug design, but there have been several applications of QSAR to the taste properties of molecules, particularly involving sweetness of different sets of compounds. QSAR studies of five families of sweet-tasting molecules have been investigated extensively.

These are sucrose, guanidine, isovanillyl, sulfamate and amino-succinamic acid derivatives with their known RS (sweetness relative to sucrose) values.^{9–5} Drew *et al.* successfully used molecular descriptors and energies derived via molecular field analysis (MFA) for the computational studies of sucrose and guanidine derivatives.⁹ They developed QSAR models for isovanillyl derivatives by applying molecular field analysis and physicochemical parameters selected by using the genetic algorithm method.¹⁰ Drew *et al.* also developed QSAR models for sulfamate derivatives by using molecular field analysis followed by selection of relevant grid points by the genetic algorithm method to distinguish sweet, sweet-bitter and bitter molecules.¹¹

In this paper, a QSAR study of 31 sucrose derivatives and 30 guanidine derivatives has been performed. Sucrose and guanidine derivatives constitute a large class of sweet-tasting compounds in which there is a high degree of structural similarity and a wide range of sweetness. The QSAR study of sucrose and guanidine derivatives has been made with the help of eight parameters, *viz.* electron affinity, ionization potential, electrophilicity index, total energy, heat of formation, steric energy, molar refractivity and solvent accessible surface area. Recently, these parameters have been found useful in the QSAR study of various compounds.^{16–21}

Drew *et al.* used molecular descriptors and energies for the QSAR study of sucrose and guanidine derivatives.⁹ QSAR models for sucrose derivatives obtained by them have good predictive quality with r^2 values in the range of 0.90 and rCV^2 values in the range of 0.85. The QSAR models for sucrose derivatives developed by us also have good predictive ability with r^2 values in the range of 0.85 and rCV^2 values in the range of 0.80. However, a

* To whom correspondence should be addressed. E-mail: madil.khan207@gmail.com

mono-parametric QSAR model using ionization potential is obtained with r^2 0.727709 and rCV^2 0.713409. In the case of guanidine derivatives, the predictive ability of the QSAR models developed by Drew *et al.* were not as high as for the sucrose derivatives with r^2 values around 0.70 and rCV^2 values between 0.50 and 0.60. In our study, the statistical measures of the regression models for guanidine derivatives are also not as high as for the sucrose derivatives, however, a mono-parametric QSAR model using molar refractivity is obtained with r^2 0.744515 and rCV^2 0.703491.

2. Materials and Method

Thirty one derivatives of sucrose and thirty derivatives of guanidine, which have been taken from the literature, were used as study material. These are listed in Table 1 and Table 5 along with their observed RS (sweetness relative to sucrose) values. The QSAR studies of both sets of derivatives have been made with the help of eight descriptors, *viz.* electron affinity, ionization potential, electrophilicity index, total energy, heat of formation, steric energy, molar refractivity and solvent accessible surface area. The geometry optimization of all the derivatives and evaluation of values of the descriptors have been done with the help of CAChe Pro from Fujitsu software. The density functional theory^{22–25} based descriptors, *viz.* electron affinity, ionization potential, electrophilicity index and total energy have been calculated by using the DFT-B88-PW91 GGA functional with the DZVP basis sets. The values of heat of formation and steric energy have been obtained by using the PM3 method²⁶ and the solvent accessible surface area (SASA) was calculated at an optimized geometry in water. The water geometry was from an optimization by using MOPAC with PM3 parameters and the Conductor like Screening Model (COSMO).²⁷ The molar refractivity was calculated by the atom typing scheme of Ghose and Crippen.²⁸ The Project Leader program associated with CAChe Pro was used for multi-linear regression (MLR) analysis. The statistical parameters were calculated by Smith's Statistical Package (version 2.80). The values of the descriptors were evaluated by solving the relevant equations given below.

According to the Koopman's theorem, the ionization potential is simply the eigenvalue of the highest occupied molecular orbital (HOMO) with change of sign and the electron affinity is the eigenvalue of the lowest unoccupied molecular orbital (LUMO) with change of sign.²⁹

Parr *et al.* introduced the electrophilicity index (ω) in terms of the chemical potential and hardness.³⁰ The operational definition of the electrophilicity index may be written as,

$$\omega = \mu^2/2\eta.$$

The total energy (TE) of a molecular system is the sum of the total electronic energy (E_{ee}) and the energy of internuclear repulsion (E_{nr}).³¹

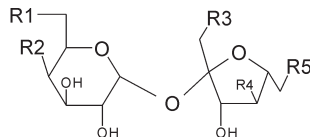
$$TE = E_{ee} + E_{nr}.$$

The total electronic energy of the system is given by

$$E_{ee} = \frac{1}{2} P(H + F),$$

where P is the density matrix, H is the one-electron matrix and F is the Fock matrix. The Hartree-Fock method is an *ab initio* method based on averaged electron-electron interactions. The Hartree-Fock method is generally derived by assuming a specific form of the solution to the quantum mechanical equation as expressed in the electronic Schrödinger equation. This solution leads to a set of coupled homogeneous equations called the Hartree-Fock equations. The Hartree-Fock equations can be

Table 1 Structure of the first set of compounds, i.e. 31 sucrose derivatives, along with their logRS value.



No.	R1	R2	R3	R4	R5	logRS
1	OH	OH	OH	OH	OH	-0.6989
2	OH	OH	OH	Cl	OH	0.3010
3	Cl	Cl	OH	OH	Cl	0.6020
4	OH	Cl	OH	OH	OH	0.6989
5	OH	OH	OH	Cl	Cl	0.6989
6	OH	OH	Cl	OH	OH	1.3010
7	OH	OH	OH	OH	Cl	1.3010
8	Cl	OH	Cl	OH	Cl	1.3979
9	OH	OH	Cl	Cl	OH	1.4771
10	OH	F	F	OH	F	1.6020
11	OH	Cl	OH	OH	Cl	1.6989
12	OH	OH	Cl	OH	Cl	1.8808
13	OH	OH	Cl	Cl	Cl	2.0000
14	OH	Cl	Cl	OH	OH	2.0791
15	OH	I	I	OH	I	2.0791
16	OH	Cl	Cl	H	Cl	2.1760
17	OH	Cl	OH	Cl	Cl	2.2041
18	Cl	Cl	Cl	OH	Cl	2.3010
19	OH	Cl	Cl	Cl	OH	2.3424
20	OH	Cl	Cl	OMe	Cl	2.4771
21	OH	Br	Cl	OH	Cl	2.5740
22	H	Cl	Cl	OH	Cl	2.6020
23	OMe	Cl	Cl	OH	Cl	2.6989
24	OH	Cl	Cl	OH	Cl	2.8129
25	OH	Cl	Br	OH	Br	2.9030
26	OH	Br	Br	OH	Br	2.9030
27	OH	Cl	Cl	F	Cl	3.0000
28	OH	Cl	Cl	Cl	Cl	3.3424
29	OH	Cl	Cl	Br	Cl	3.4771
30	OH	Cl	Cl	I	Cl	3.8750
31	OH	Br	Br	Br	Br	3.8750

written in matrix form which is known as the Fock matrix. The total energy of the molecular system is a function of the positions of the atoms and one-particle wave functions. A density matrix is defined over the occupied orbitals and can be used along with the one- and two-electron integrals of the atomic basis in an appropriate representation of the Fock matrix. In a Hartree-Fock solution procedure, the molecular orbital coefficients are used to compute the density matrix, which in turn is used to construct the Fock matrix from the list of atomic orbital two-electron integrals.

The heat of formation is defined as:

$$\Delta H_f = E_{\text{elect}} + E_{\text{nuc}} - E_{\text{isol}} + E_{\text{atom}},$$

where E_{elect} is the electronic energy, E_{nuc} is the nuclear-nuclear repulsion energy, E_{isol} is the energy required to strip all the valence electrons of all the atoms in the system and E_{atom} is the total heat of atomization of all the atoms in the system.³²

The steric energy of a molecule is the sum of the molecular mechanics potential energies calculated for the bonds, bond angles, dihedral angles, nonbonded atoms and so forth. It is specific to mechanics and depends upon the force-field used.³³

The solvent accessibility surface area (SASA) is the surface area

Table 2 Values of the descriptors for the 31 sucrose derivatives.

No.	EA	IP	ω	MR	SASA	TE	ΔH_f	SE
1	0.113	5.718	1.516	68.774	132.116	–1297.997	–477.195	130.743
2	0.557	5.592	1.878	71.825	135.699	–1682.344	–440.496	117.801
3	0.502	5.619	1.830	77.927	152.919	–2451.019	–363.996	114.573
4	0.421	5.596	1.749	71.825	137.370	–1682.335	–441.138	121.853
5	0.688	5.586	2.009	74.876	140.247	–2066.684	–400.947	116.732
6	0.508	5.723	1.861	71.825	135.589	–1682.336	–438.212	122.617
7	–0.175	5.614	1.277	71.825	133.584	–1682.342	–438.670	126.687
8	0.329	5.871	1.734	77.927	152.595	–2451.015	–359.039	114.619
9	0.736	5.888	2.129	74.876	142.065	–2066.680	–401.520	114.594
10	–0.321	5.976	1.269	63.698	131.214	–1370.073	–479.375	111.941
11	0.337	5.728	1.706	74.876	139.270	–2066.688	–402.632	117.804
12	0.417	5.886	1.816	74.876	139.818	–2066.690	–399.686	118.920
13	0.865	5.891	2.270	77.927	146.285	–2451.029	–362.811	108.918
14	0.559	5.790	1.926	74.876	140.650	–2066.681	–402.218	113.527
15	1.430	5.852	2.997	102.919	183.213	–21831.257	–269.022	101.888
16	0.453	6.018	1.881	76.830	145.349	–2375.806	–323.533	91.919
17	0.792	5.833	2.177	77.927	145.483	–2451.031	–365.015	107.813
18	0.597	5.989	2.011	80.978	157.941	–2835.362	–324.988	110.836
19	0.737	5.911	2.135	77.927	147.355	–2451.027	–365.530	105.431
20	0.387	5.923	1.798	82.678	152.621	–2490.338	–358.418	107.722
21	1.286	6.009	2.817	80.911	148.479	–4564.716	–353.596	108.281
22	0.391	5.924	1.802	76.383	146.250	–2375.798	–323.364	107.853
23	0.395	5.657	1.740	82.678	159.036	–2490.323	–356.119	114.209
24	0.494	6.020	1.919	77.927	145.051	–2451.034	–363.678	109.790
25	1.110	5.973	2.580	83.894	149.738	–6678.402	–341.980	108.112
26	1.250	5.920	2.751	86.878	156.431	–8792.080	–331.894	106.298
27	0.543	6.115	1.989	76.235	143.952	–2475.061	–364.739	99.798
28	0.912	6.105	2.370	80.978	151.349	–2835.376	–326.350	99.799
29	1.694	6.031	3.441	83.962	155.358	–4949.061	–315.982	98.052
30	1.713	6.054	3.473	89.309	162.623	–9295.451	–296.468	97.150
31	1.618	5.974	3.308	92.912	166.505	–11290.109	–284.652	95.312

EA = electron affinity (eV); IP = ionization potential (eV); ω = electrophilicity index; MR = molar refractivity; SASA = solvent accessible surface area; TE = total energy (Hartree); ΔH_f = heat of formation (kcal/mol); SE = steric energy (kcal/mol).

of a biomolecule that is accessible to a solvent and is usually quoted in square angstroms. Lee and Richards first described the solvent accessible surface area (SASA) of a molecular surface.³⁴ SASA is typically calculated by using the 'rolling ball' algorithm developed by Sharke and Rupley.³⁵ This algorithm uses a sphere of solvent of a particular radius to probe the surface of the molecule. The choice of the probe radius does have an effect on the observed surface area, as using a smaller probe radius detects more surface details and therefore reports a larger surface. A typical value is 1.4 Å, which approximates the radius of a water molecule.

Molar refractivity is calculated by the Lorenz-Lorentz formula³⁶

$$MR = \frac{n^2 - 1}{n^2 + 1} \times \frac{M}{\rho},$$

where M is the molecular weight, n is the refractive index and ρ is the density. For a radiation of infinite wavelength, molar refractivity represents the real volume of the molecules.

3. Results and Discussion

3.1. QSAR Study of the First Set of Compounds Containing Sucrose Derivatives

Thirty one derivatives of sucrose are given in Table 1 along with their observed sweetness in terms of logRS values. The values of the eight descriptors of the compounds, which have

been calculated, are given in Table 2. For the development of QSAR models multi-linear regression (MLR) analysis has been performed by using different combinations of the descriptors. Compound numbers 1 and 23 are outliers in the MLR analysis, i.e. removing these two compounds from the data set greatly increases the predictive ability of the QSAR models. The MLR analysis has indicated that the sweetness of sucrose can be successfully modelled even in mono-parametric regression by using ionization potential as a descriptor. This mono-parametric QSAR model is obtained from the following regression equation,

$$SUCROSE_{PS1} = 5.08109 \times IP - 27.7204.$$

$r^2 = 0.727709$, $rCV^2 = 0.713409$, $SE = 0.1177$, $SEE = 0.5028$, $t\text{-value} = 8.4963$, $P\text{-value} = 0$, $DOF = 0.7177$, $n = 29$.

In the above regression equation, r^2 is the squared correlation coefficient, rCV^2 is the cross-validation coefficient, S.E is the standard error, SEE is the standard error of the estimate, DOF is the degrees of freedom and n is the number of data points (compounds). The ionization potential appears to be an important descriptor for this set of sucrose derivatives. The trend of observed sweetness and predicted sweetness obtained from $SUCROSE_{PS1}$ is shown in Fig. 1.

The addition of the electrophilicity index (ω) in the above mono-parametric model yields a model with dramatically improved predictability. The resulting bi-parametric QSAR model is obtained from the following regression equation,

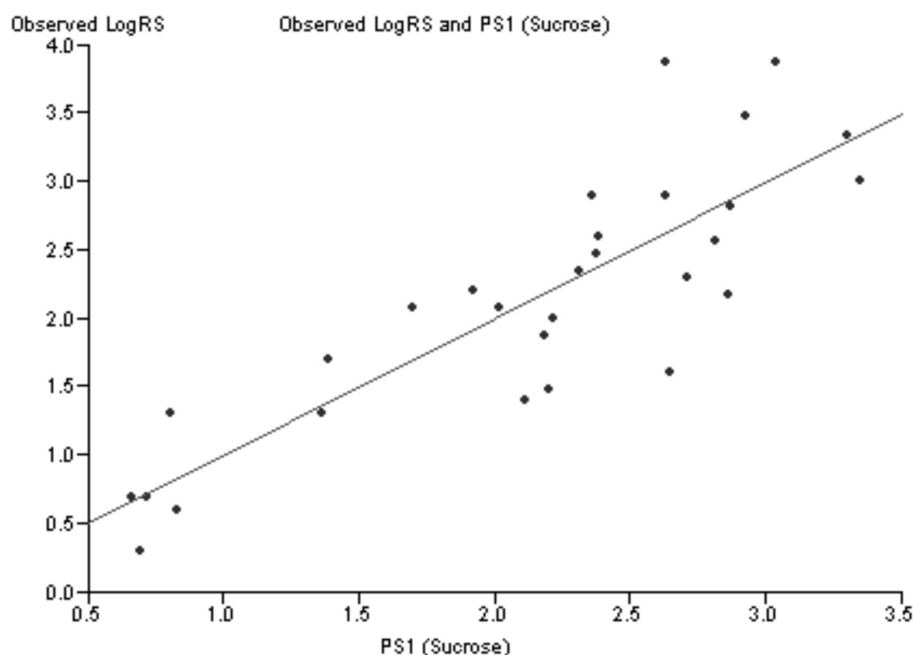


Figure 1 Trend of observed sweetness (logRS) and predicted sweetness (obtained from $SUCROSE_{PS1}$) of the sucrose derivatives.

$$SUCROSE_{PS2} = 4.07088 \times IP + 0.616562 \times \omega - 23.1215.$$

$r^2 = 0.842159$, $rCV^2 = 0.805992$, $SE = 0.0833$, $SEE = 0.3829$, $t\text{-value} = 12.0028$, $P\text{-value} = 0$, $DOF = 0.8363$, $n = 29$.

The trend of observed sweetness and predicted sweetness obtained from $SUCROSE_{PS2}$ is shown in Fig. 2.

The best tri-parametric QSAR model is developed by using the descriptors ionization potential, molar refractivity and solvent accessibility surface area (SASA). This tri-parametric QSAR model is obtained from the following regression equation,

$$SUCROSE_{PS3} = 4.63976 \times IP + 0.138893 \times MR - 0.0712653 \times SASA - 25.5542.$$

$r^2 = 0.865448$, $rCV^2 = 0.816901$, $SE = 0.0759$, $SEE = 0.3536$, $t\text{-value} = 13.1759$, $P\text{-value} = 0$, $DOF = 0.8604$, $n = 29$.

The trend of observed sweetness and predicted sweetness obtained from $SUCROSE_{PS3}$ is shown in Fig. 3.

From the values of the squared correlation coefficient (r^2), cross-validation coefficient (rCV^2) and other statistical parameters for the above three QSAR models, it is clear that the predictive power of all models is high. Among these three QSAR models the tri-parametric model, i.e. $SUCROSE_{PS3}$, is the best. It can be used to find the sweetness value of any new derivative of sucrose. The predicted logRS values, for sucrose derivatives of this set, obtained from above three QSAR models are listed in Table 3 along with their observed logRS values. A correlation summary of the best three QSAR models for sucrose derivatives of this set is presented in Table 4.

3.2. QSAR Study of the Second Set of Compounds Containing Guanidine Derivatives.

Thirty derivatives of guanidine are given in Table 5 along with their observed sweetness in terms of logRS values. The values of the eight descriptors of the compounds, which have been calcu-

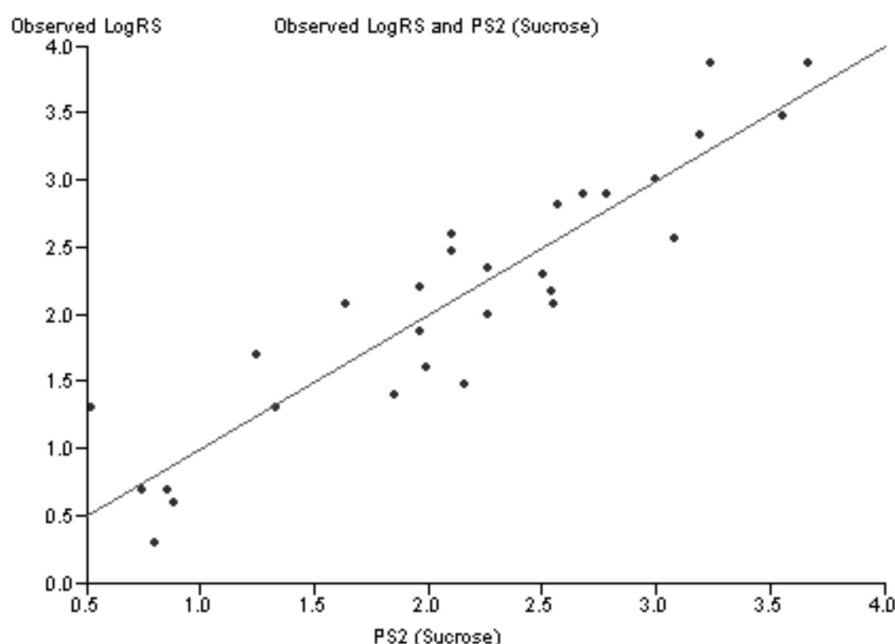


Figure 2 Trend of observed sweetness (logRS) and predicted sweetness (obtained from $SUCROSE_{PS2}$) of the sucrose derivatives.

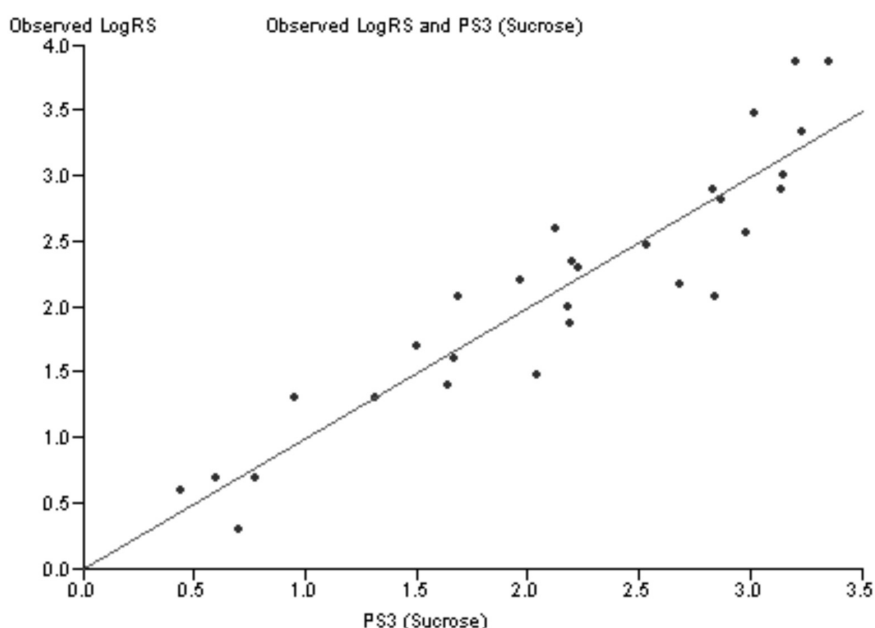


Figure 3 Trend of observed sweetness (logRS) and predicted sweetness (obtained from $SUCROSE$ PS3) of the sucrose derivatives.

lated, are given in Table 6. Multi-linear regression (MLR) analysis has been performed by using different combinations of the descriptors. Compound numbers 3, 14 and 24 are outliers in the MLR analysis, i.e. removing these three compounds from the data set greatly increases the predictive ability of the QSAR models. From the MLR analysis, it was found that the sweetness of guanidine can be successfully modelled by mono-parametric regression by using molar refractivity as the descriptor. This best mono-parametric QSAR model is obtained from the following regression equation,

$$GUANIDINEPS1 = 0.0662468 \times MR - 1.94682.$$

$r^2 = 0.744515$, $rCV^2 = 0.703491$, $SE = 0.1171$, $SEE = 0.2956$, $t\text{-value} = 8.5371$, $P\text{-value} = 0$, $DOF = 0.7344$, $n = 27$.

Molar refractivity appears to be an important descriptor for this set of guanidine derivatives. The trend of observed sweetness and predicted sweetness obtained from $GUANIDINE$ PS1 is shown in Fig. 4.

The best bi-parametric QSAR model for this set of derivatives is obtained by the addition of the solvent accessibility surface area (SASA) in the above mono-parametric model. This best bi-parametric QSAR model is obtained from the following regression equation,

$$GUANIDINEPS2 = 0.0541185 \times MR + 0.0133308 \times SASA - 2.88913.$$

$r^2 = 0.763869$, $rCV^2 = 0.694028$, $SE = 0.1112$, $SEE = 0.2842$, $t\text{-value} = 8.9917$, $P\text{-value} = 0$, $DOF = 0.7544$, $n = 27$.

The trend of observed sweetness and predicted sweetness obtained from $GUANIDINE$ PS2 is shown in Fig. 5.

The best tri-parametric QSAR model is developed by using the

Table 3 Observed and predicted sweetness (in terms of logRS) of sucrose derivatives of the first set.

No.	Observed logRS	Predicted logRS		
		$SUCROSE$ PS1	$SUCROSE$ PS2	$SUCROSE$ PS3
2	0.3010	0.693	0.801	0.697
3	0.6020	0.830	0.881	0.442
4	0.6989	0.713	0.738	0.596
5	0.6989	0.663	0.857	0.769
6	1.3010	1.359	1.324	1.312
7	1.3010	0.805	0.520	0.950
8	1.3979	2.111	1.848	1.635
9	1.4771	2.197	2.160	2.040
10	1.6020	2.644	1.988	1.669
11	1.6989	1.384	1.248	1.497
12	1.8808	2.187	1.959	2.191
13	2.0000	2.212	2.260	2.177
14	2.0791	1.699	1.636	1.686
15	2.0791	2.014	2.549	2.836
16	2.1760	2.858	2.537	2.681
17	2.2041	1.918	1.966	1.965
18	2.3010	2.710	2.499	2.225
19	2.3424	2.314	2.258	2.194
20	2.4771	2.375	2.099	2.534
21	2.5740	2.812	3.077	2.983
22	2.6020	2.380	2.105	2.118
24	2.8129	2.868	2.568	2.864
25	2.9030	2.629	2.785	3.140
26	2.9030	2.360	2.674	2.832
27	3.0000	3.350	2.998	3.148
28	3.3424	3.300	3.192	3.233
29	3.4771	2.924	3.552	3.018
30	3.8750	3.041	3.665	3.350
31	3.8750	2.634	3.238	3.203

Table 4 Correlation summary of the best three QSAR models for sucrose derivatives of the first set.

QSAR Model	r^2	rCV^2	SE	SEE	t-value	P-value	DOF	VC	Variable used
$SUCROSE$ PS1	0.7277	0.7134	0.1177	0.5028	8.4963	0	0.7177	1	IP
$SUCROSE$ PS2	0.8422	0.8060	0.0833	0.3829	12.0028	0	0.8363	2	IP, ω
$SUCROSE$ PS3	0.8654	0.8169	0.0759	0.3536	13.1759	0	0.8604	3	IP, MR, SASA

r^2 = Squared correlation coefficient; rCV^2 = cross-validation coefficient; SE = standard error; SEE = standard error of estimate; DOF = degrees of freedom; VC = variable count; IP = ionization potential; ω = electrophilicity index; MR = molar refractivity; SASA = solvent accessible surface area.

Table 5 Structure of the second set of compounds, i.e. 30 guanidine derivatives, along with their logRS value.

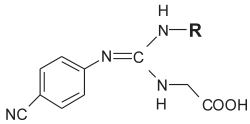
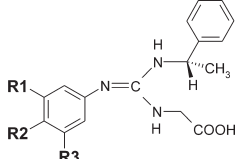
							
No.	R	logRS	No.	R1	R2	R3	logRS
1	c-C ₉ H ₁₇	5.3010	20	H	H	H	3.6989
2	c-C ₁₀ H ₁₉	5.1760	21	Br	H	H	4.3979
3	c-C ₇ H ₁₃	4.7781	22	CH ₃	H	H	4.0791
4	1-naphthyl	4.7781	23	CF ₃	H	H	3.8750
5	(S)CH(CH ₃)-c-C ₆ H ₁₁	4.6989	24	CN	H	H	3.7403
6	CH ₂ -c-C ₆ H ₁₁	4.5440	25	Cl	H	Cl	5.0791
7	CH ₂ -C ₆ H ₅	4.4771	26	CH ₃	H	CH ₃	4.4771
8	(S)CH(CH ₃)C ₆ H ₅	4.4471	27	F	H	F	4.1760
9	CH ₂ -1-adamantyl	4.3617	28	CH ₃	CN	H	4.6989
10	N(CH ₃)C ₆ H ₅	4.2552	29	CH ₃	CN	CH ₃	4.6989
11	c-C ₆ H ₁₁	4.0791	30	Cl	Cl	Cl	4.5440
12	C ₆ H ₄ (3-Cl)	4.0000					
13	C ₆ H ₄ (3-CH ₃)	3.9542					
14	CH ₂ CH ₂ C ₆ H ₅	3.9294					
15	C ₆ H ₄ (4-CH ₃)	3.8450					
16	(CH ₂) ₅ CH ₃	3.7781					
17	C ₆ H ₄ (2-CH ₃)	3.6989					
18	C ₆ H ₅	3.6020					
19	CH ₂ CH ₃	2.5440					

Table 6 Values of the descriptors for the 30 guanidine derivatives.

No.	EA	IP	ω	MR	SASA	TE	ΔH _f	SE
1	1.973	5.558	3.955	101.374	157.521	-1109.092	-43.192	18.240
2	1.854	5.542	3.709	105.975	167.918	-1148.399	-44.277	26.640
3	1.870	5.571	3.741	92.172	153.404	-1030.490	-32.742	11.187
4	2.398	5.063	5.221	101.300	162.541	-1141.178	35.575	-14.778
5	2.008	5.050	4.094	96.721	160.001	-1069.809	-42.472	6.310
6	1.870	5.590	3.741	92.302	156.216	-1030.501	-37.969	4.025
7	1.911	5.645	3.822	91.170	153.148	-1026.875	12.422	-13.815
8	1.929	5.642	3.859	95.588	159.852	-1066.190	8.145	-13.106
9	1.893	5.396	3.793	90.245	170.061	-1184.047	-1.736	1846.493
10	1.994	5.228	4.032	87.751	157.994	-1082.123	40.636	37.088
11	1.866	5.580	3.733	87.571	148.647	-991.184	-33.345	-0.102
12	2.290	5.772	4.667	89.655	159.143	-1447.119	10.926	-12.196
13	1.968	5.567	3.944	89.891	153.565	-1026.857	10.363	-6.659
14	1.923	5.659	3.847	95.925	162.711	-1066.191	8.055	-9.672
15	1.967	5.343	3.956	89.891	151.722	-1026.855	7.885	-8.112
16	1.890	5.610	3.780	89.633	160.091	-992.382	-40.661	-2.127
17	2.094	5.532	4.229	89.891	152.975	-1026.861	8.513	-9.663
18	2.329	5.203	4.934	84.850	146.521	-987.543	19.025	-6.709
19	2.056	5.082	4.210	71.306	128.274	-835.135	-21.543	-4.493
20	1.236	5.116	2.599	89.851	142.183	-973.932	-28.661	-17.011
21	1.316	5.280	2.743	97.474	159.716	-3547.196	-21.138	-18.130
22	1.258	5.052	2.623	94.892	146.735	-1013.250	-38.250	-17.897
23	1.705	5.509	3.419	95.824	159.492	-1311.043	-186.177	-10.797
24	2.152	5.617	4.355	95.588	157.129	-1066.183	7.691	-15.141
25	1.425	5.510	2.943	99.460	164.349	-1893.085	-41.061	-21.316
26	1.234	4.971	2.576	99.933	153.728	-1052.571	-46.386	-17.004
27	1.403	5.471	2.904	90.283	148.316	-1172.457	-114.320	-23.883
28	2.012	5.560	4.040	100.630	157.796	-1105.504	-1.636	-12.996
29	1.962	5.471	3.936	105.671	163.615	-1144.819	-10.714	-11.368
30	1.568	5.400	3.168	104.265	171.291	-2352.648	-45.600	-13.672

EA = electron affinity (eV); IP = ionization potential (eV); ω = electrophilicity index; MR = molar refractivity; SASA = solvent accessible surface area; TE = total energy (Hartree); ΔH_f = heat of formation (kcal/mol); SE = steric energy (kcal/mol).

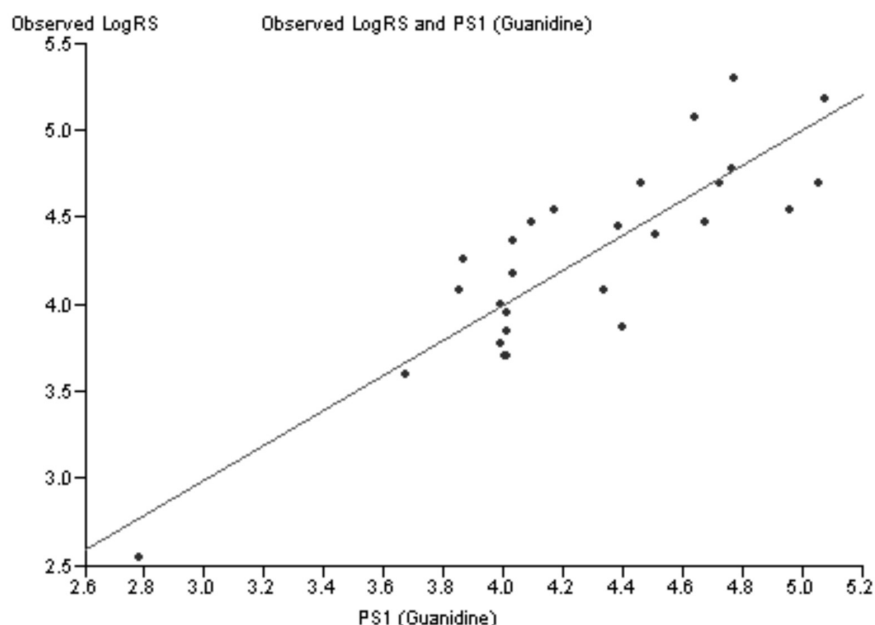


Figure 4 Trend of observed sweetness (logRS) and predicted sweetness (obtained from $^{GUANIDINE}PS1$) of the guanidine derivatives.

descriptors ionization potential, molar refractivity and heat of formation. This tri-parametric QSAR model is obtained from the following regression equation,

$$^{GUANIDINE}PS3 = 0.33132 \times IP + 0.0669111 \times MR + 0.00167667 \times \Delta H_f - 3.75749.$$

$r^2 = 0.777595$, $rCV^2 = 0.616007$, $SE = 0.1070$, $SEE = 0.2758$, $t\text{-value} = 9.3485$, $P\text{-value} = 0$, $DOF = 0.7687$, $n = 27$.

The trend of observed sweetness and predicted sweetness obtained from $^{GUANIDINE}PS3$ is shown in Fig. 6.

From the values of the squared correlation coefficient (r^2), cross-validation coefficient (rCV^2) and other statistical parameters for the above three QSAR models, it is clear that the predictive power of all models is high. Among these three QSAR models the tri-parametric model, i.e. $^{GUANIDINE}PS3$, is the best which can be used to find the sweetness value of any new derivative of guanidine. The predicted logRS values, for guanidine derivatives of this set, obtained from above three QSAR models are listed in

Table 7 along with their observed logRS values. A correlation summary of the best three QSAR models for guanidine derivatives of this set is presented in Table 8.

4. Conclusions

A reference to Table 4 indicates that ionization potential appears an important descriptor for sucrose derivatives. Ionization potential in combination with electrophilicity index, molar refractivity and solvent accessibility surface area provide better results. The best combination of descriptors obtained for predicting the sweetness value of sucrose derivatives is ionization potential, molar refractivity and solvent accessibility surface area.

Reference to Table 8 indicates that molar refractivity appears an important descriptor for guanidine derivatives. Molar refractivity in combination with solvent accessibility surface area, ionization potential and heat of formation provide better results. The best combination of descriptors obtained for predict-

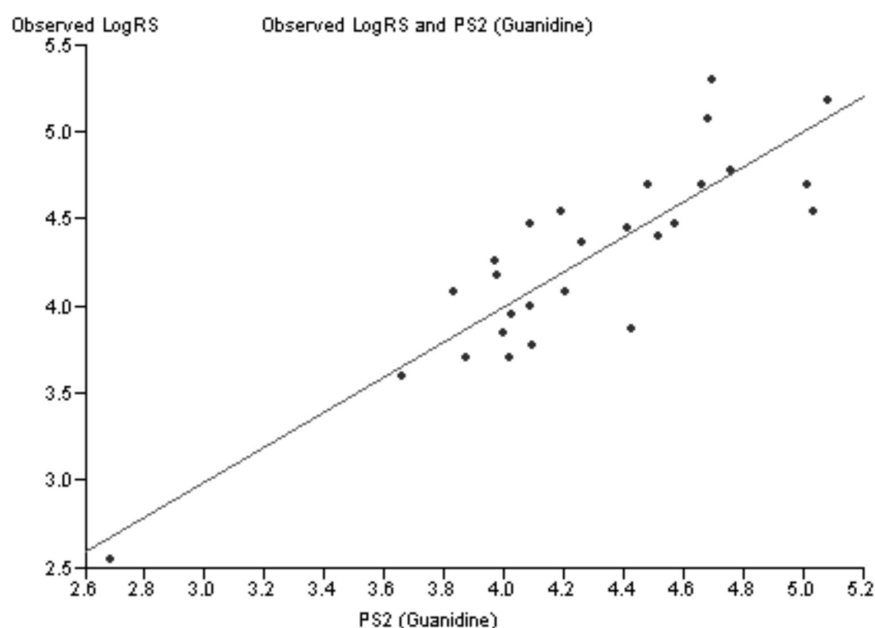


Figure 5 Trend of observed sweetness (logRS) and predicted sweetness (obtained from $^{GUANIDINE}PS2$) of the guanidine derivatives.

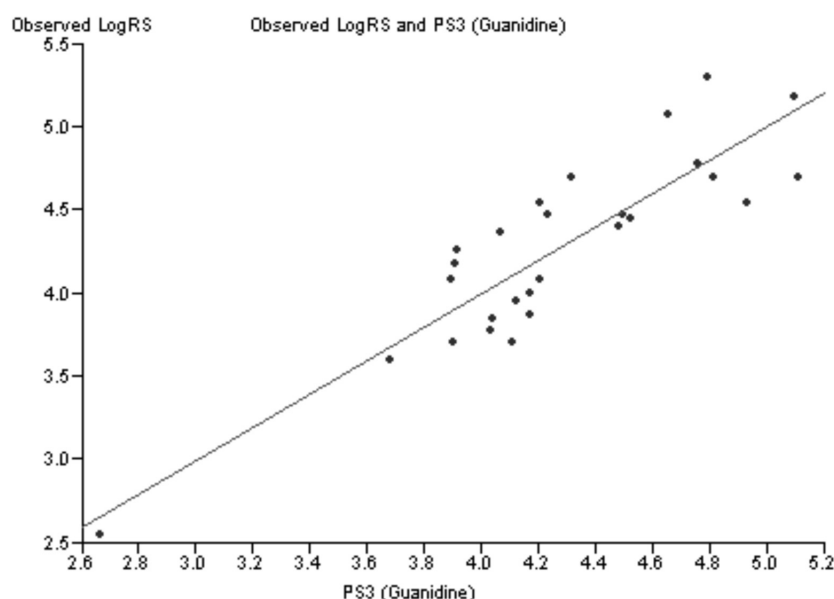


Figure 6 Trend of observed sweetness (logRS) and predicted sweetness (obtained from ^{GUANIDINE}PS3) of the guanidine derivatives.

ing the sweetness value of guanidine derivatives is ionization potential, molar refractivity and heat of formation. Molar refractivity is related not only to the volume of the molecules but also to the London dispersive forces. Thus London dispersive forces appear to play an important role in guanidine-receptor interaction.

References

- 1 B. Lindemann, *Physiol. Rev.* 2001, **76**, 718–766.
- 2 M. Kitagawa, Y. Kusakabe and H. Mius, *Biochem. Biophys. Res. Commun.* 2001, **283**, 236–242.
- 3 E. Sainz, J.N. Korley and J.F. Battey, *J. Neurochem.* 2001, **77**, 896–903.
- 4 X.D. Li, L. Staszewski, H. Xu, K. Durick, M. Zoller and E. Adler, *Proc. Natl. Acad. Sci.* 2002, **99**(7), 4692–4696.
- 5 R.S. Shallenberger and T.E. Acree, *Nature* 1967, **216**, 480–482.
- 6 L.B. Kier, *J. Pharm. Sci.* 1972, **61**, 1394–1397.
- 7 C. Nofre and J.M. Tinti, *Food Chem.* 1996, **56**, 263–274.
- 8 C. Nofre, J.M. Tinti and D. Glaser, *Chem. Senses.* 1996, **21**, 747–762.
- 9 J.S. Barker, C.K. Hattotuwigama and M.G.B. Drew, *Pure Appl. Chem.*, 2002, **74**(7), 1207–1217.
- 10 A. Bassoli, M.G.B. Drew, C.K. Hattotuwigama, L. Merlini, G. Morini and G.R.H. Wilden, *Quant. Struct.-Act. Relat.* 2001, **20**, 3–16.
- 11 M.G.B. Drew, G.R.H. Wilden, W.J. Spillane, R.M. Walsh, C.A. Ryder and J.M. Simmie, *J. Agric. Food Chem.* 1998, **46**, 3016–3026.
- 12 A. Bassoli, L. Merlini and G. Morini, *Pure Appl. Chem.*, 2002, **74**(7), 1181–1187.
- 13 W.J. Spillane, C.M. Coyle, B.G. Feeney and E.F. Thompson, *J. Agric. Food Chem.*, 2009, **57**(12), 5486–5493.
- 14 L. Tarkoa, I. Lupescub and D.C. Groposilab, *ARKIVOC*, 2006, (xiii), 22–40.
- 15 L. Tarkoa, I. Lupescub and D.C. Groposilab, *ARKIVOC*, 2005, (x), 254–271.

Table 7 Observed and predicted sweetness (in terms of logRS) of guanidine derivatives of the second set.

No.	Observed logRS	Predicted logRS		
		^{GUANIDINE} PS1	^{GUANIDINE} PS2	^{GUANIDINE} PS3
1	5.3010	4.769	4.697	4.795
2	5.1760	5.074	5.085	5.095
4	4.7781	4.764	4.760	4.758
5	4.6989	4.461	4.478	4.316
6	4.5440	4.168	4.189	4.207
7	4.4771	4.093	4.086	4.234
8	4.4471	4.386	4.415	4.521
9	4.3617	4.032	4.262	4.066
10	4.2552	3.866	3.966	3.914
11	4.0791	3.854	3.832	3.895
12	4.0000	3.993	4.084	4.172
13	3.9542	4.008	4.023	4.119
15	3.8450	4.008	3.998	4.041
16	3.7781	3.991	4.096	4.030
17	3.6989	4.008	4.015	4.104
18	3.6020	3.674	3.656	3.676
19	2.5440	2.777	2.680	2.661
20	3.6989	4.006	3.869	3.902
21	4.3979	4.511	4.515	4.479
22	4.0791	4.339	4.202	4.202
23	3.8750	4.401	4.423	4.167
25	5.0791	4.642	4.684	4.654
26	4.4771	4.673	4.568	4.498
27	4.1760	4.034	3.974	3.904
28	4.6989	4.720	4.660	4.815
29	4.6989	5.054	5.011	5.108
30	4.5440	4.960	5.037	4.932

Table 8 Correlation summary of the best three QSAR models for the guanidine derivatives of the second set.

QSAR Model	r ²	rCV ²	SE	SEE	t-value	P-value	DOF	VC	Variable used
^{GUANIDINE} PS1	0.7445	0.7035	0.1171	0.2956	8.5371	0	0.7344	1	MR
^{GUANIDINE} PS2	0.7639	0.6940	0.1112	0.2842	8.9917	0	0.7544	2	MR, SASA
^{GUANIDINE} PS3	0.7776	0.6160	0.1070	0.2758	9.3485	0	0.7687	3	IP, MR, ΔH _f

r² = squared correlation coefficient; rCV² = cross-validation coefficient; SE = standard error; SEE = standard error of estimate; DOF = degrees of freedom; VC = variable count; MR = molar refractivity; SASA = solvent accessible surface area; IP = ionization potential; ΔH_f = heat of formation.

- 16 V.K. Sahu, A.K.R. Khan, R.K. Singh and P.P. Singh, *Int. Jour. Quant. Chem.* 2009, **109**(6), 1243–1254.
- 17 F.A. Pasha, S.J. Cho, Y.Beg and Y.B. Tripathi, *Med. Chem. Res.* 2007, **16**(7–9), 408–417.
- 18 H.K. Srivastava, F.A. Pasha, S.K. Mishra and P.P. Singh, *Med. Chem. Res.* 2009, **18**, 455–466.
- 19 A.K.R. Khan, S.K. Mishra, S.A. Khan and M. Ansari, *Med. Chem. Res.*, 2011, **21**(9), 2153–2161.
- 20 R.K. Singh, A.K.R. Khan, V.K. Sahu and P.P. Singh, *Int. Jour. Quant. Chem.* 2009, **109**(2), 185–195.
- 21 V.K. Sahu and R. K. Singh, *CLEAN-Soil, Air, Water*, 2009, **37**(11), 850–857.
- 22 R.G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, 1989, New York.
- 23 W. Kohn, A.D. Becke, R.G. Parr, *J. Phys. Chem.* 1996, **100**, 12974–12980.
- 24 W. Kohn, *Rev. Mod. Phys.* 1999, **71**(5), 1253–1266.
- 25 K. Capelle, *Braz. Jour. of Phys.* 2006, **36**(4A), 1318–1343.
- 26 J.J.P. Stewart, *J. Comp. Chem.*, 1989, **10**(2), 209–220.
- 27 A. Klamt and G. Schuurmann, *J. Chem. Soc. Perkin Trans.* 1993, **2**, 799–805.
- 28 A.K. Ghose, A. Pritchett and G.M. Crippen, *J. Comput. Chem.* 1988, **9**, 80–90.
- 29 R.G. Parr, R.A. Donnelly, M. Levy and W.E. Palke, *J. Chem. Phys.* 1978, **68**, 3801–3807.
- 30 R.G. Parr, L.V. Szentpaly and S. Liu, *J. Am. Chem. Soc.* 1999, **121**, 1922–1924.
- 31 R.G. Parr and W. Young, *Annu. Rev. Phys. Chem.* 1995, **46**, 701–728.
- 32 N. Bodor, Z. Gabanyi and C.K. Wong, *J. Am. Chem. Soc.* 1989, **111**, 3783–3786.
- 33 T. Clark, *A Handbook of Computational Chemistry*, John Wiley and Sons, 1985, New York.
- 34 B. Lee and F.M. Richards, *J. Mol. Biol.* 1971, **55**(3), 379–400.
- 35 A. Shrake and J.A. Rupley, *J. Mol. Biol.* 1973, **79**(2), 351–371.
- 36 R.J. Padron, A. Carrasco and R.F. Pellon, *J. Pharm. Pharmaceut. Sci.* 2002, **5**, 258–266.