

See discussions, stats, and author profiles for this publication at:  
<https://www.researchgate.net/publication/291953313>

# A New QSPR Study on Relative Sweetness

Article · January 2016

DOI: 10.4018/IJQSPR.2016010104

---

CITATIONS

4

---

READS

102

3 authors, including:



**Cristian Rojas Villa**

National University of La Plata

57 PUBLICATIONS 49 CITATIONS

SEE PROFILE



**Piercosimo Tripaldi**

Universidad del Azuay (UDA)

63 PUBLICATIONS 48 CITATIONS

SEE PROFILE

All content following this page was uploaded by [Cristian Rojas Villa](#) on 01 March 2016.

The user has requested enhancement of the downloaded file.

# A New QSPR Study on Relative Sweetness

Cristian Rojas, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (CCT La Plata-CONICET, UNLP), La Plata, Argentina & Decanato General de Investigaciones, Universidad del Azuay, Cuenca, Ecuador

Piercosimo Tripaldi, Laboratorio de Química-Física de Alimentos, Facultad de Ciencia y Tecnología, Universidad del Azuay, Cuenca, Ecuador

Pablo R. Duchowicz, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (CCT La Plata-CONICET, UNLP), La Plata, Argentina

## ABSTRACT

The aim of this work was to develop predictive structure-property relationships (QSPR) of natural and synthetic sweeteners in order to predict and model relative sweetness (RS). The data set was composed of 233 sweeteners collected from diverse sources in the literature, which was divided into training (163) and test (70) molecules according to a procedure based on k-means cluster analysis. A total of 3763 non-conformational Dragon molecular descriptors were calculated which were simultaneously analyzed through multivariable linear regression analysis coupled with the replacement method variable subset selection technique. The established six-parameter model was validated through the cross-validation techniques, together with Y-randomization and applicability domain analysis. The results for the training set and the test set showed that the non-conformational descriptors offer relevant information for modeling the RS of a compound. Thus, this model can be used to predict the sweetness of both un-evaluated and un-synthesized sweeteners.

## KEYWORDS

Dragon Software, k-Means Cluster Analysis, QSPR Theory, Relative Sweetness, Replacement Method. Sweeteners

## INTRODUCTION

Sweetness is one of the most important tastes of mankind and has long been sought after as a dietary constituent to produce a pleasant sensation. Since sucrose is the most commonly used sweetener, it is the standard substance employed for measuring and comparing the Relative Sweetness (RS) of other sweet substances (Singh, Khan, & Singh, 2014). RS is defined as the ratio of a standard sucrose concentration to the iso-sweet concentration of another sweetener (Bassoli et al., 2001). In other words, a standard solution of sucrose has a sweetness perception rating of 1 (or 100), and the sweetness of whatever sweetener is being evaluated is rated relative to sucrose.

The development and search for new sweeteners is complicated. On one hand, there are multiple factors that affect sweetness; e.g., solubility, stability at wide pH and temperature ranges, clean sweet taste without post-flavor effects, sweetening effect as compared to low-cost sucrose, and finally, the most important factor is the safety of human health (Belitz, Grosch, & Schieberle, 2009). On the other hand, the measurement of RS involves a high cost due to the use of trained panels for the use of the “sip and spit” taste method for comparing the RS between a sweetener and the standard. Because of all these factors, there is a clear advantage to develop quantitative-structure models in order to understand the sweetness mechanism and to use these models to develop and synthesize new potent sweeteners (Yang, Chong, Yan, & Chen, 2011; Zhong, Chong, Nie, Yan, & Yuan, 2013).

The QSAR/QSPR theory suggests that the biological activities and properties of molecules can be determined in terms of specific molecular attributes, i.e., these chemical structures can be represented theoretically although they do not offer specific details of the usually complex mechanism/path of action involved. However, the molecular structure is quantified by using a set of suitable molecular descriptors, which are numbers carrying information on the constitutional, topological, geometrical, hydrophobic, and/or electronic aspects of the chemical structure (Diudea, 2001; A. R. Katritzky, Lobanov, V. S., Karelson, M., 1995; Todeschini & Consonni, 2009; Trinajstić, 1992). Hence, the goal is to build a useful model selecting only the most representative descriptors among thousands of them. If successful, such a model can be useful to investigate and understand specific characteristics of the relationship between the molecular structure and its activity/property that is under investigation.

The first Quantitative Structure-Activity Relationship (QSAR) study of sweeteners was carried out by Edna W. Deutsch and Corwin Hansch (Deutsch & Hansch, 1966) in which they used some derivatives of 2-amino-4-nitrobenzene to model RS. They concluded that this property is well-correlated with the hydrophobicity and the Hammett constant descriptors given by the substituents. Since this pioneering study, there were several applications of the QSAR/QSPR theory to model and predict the RS for diverse sweeteners (Arnoldi, Bassoli, Merlini, & Ragg, 1991; Barker, Hattotuagama, & Drew, 2002; Bassoli et al., 2001; Bassoli, Drew, Merlini, & Morini, 2002; Drew et al., 1998; Hansch, 1970; Iwamura, 1980, 1981; Katritzky et al., 2002; Pietrzycki, 2001; Rao & Kumar, 1986; Spillane, 1983; Spillane & McGlinchey, 1981; Spillane, McGlinchey, Muirheartaigh, & Benson, 1983; Spillane et al., 2000; William J. Spillane et al., 1996; Spillane & Sheahan, 1989; van der Heijden, Brussel, & Peer, 1979; Walters, 2002; Walters, 2006; Walters & Hinds, 1994).

Recently, Rojas et al. (Rojas, Duchowicz, Pis Diez, & Tripaldi, 2016) published a multi-criteria review of QSAR/QSPR applications for the RS during the decade 2004-2014. This period of time is remarkable for the increased application of new QSAR/QSPR theories to investigate diverse data sets of sweeteners. In this review, the authors presented in a chronological way thirteen studies regarding the synthesis of new sweeteners and their QSAR/QSPR analysis, along with the development and application of novel methodologies to perform predictive QSAR/QSPR models. The use of such models to predict the RS value of new proposed potent sweeteners was also presented. In this review, the specific term of Quantitative Structure-Relative Sweetness Relationships (QSRSR) was introduced that refers to the property of relative sweetness.

The main purpose of the present work is to use a data set of 233 diverse natural and synthetic sweeteners for the development of a predictive non-conformational QSPR model for the RS values, contemplating the principles defined by the Organization for Economic Co-operation and Development (OECD) to make it applicable. In brief, the modelled property and the mathematical algorithm should be clearly defined, the model should be accompanied by a definition of its applicability domain, the goodness-of-fit and predictivity of the model should be evaluated through appropriate strategies and, eventually, a mechanistic interpretation of model descriptors should be given, if possible. Dragon molecular descriptors and the replacement method variable subset selection were used for the first time for this purpose. The predictive ability was estimated by means of the internal and external validation procedures, and the chemical information encoded in the selected descriptors was explained. If this QSPR model shows an acceptable prediction of the relative sweetness, it may be a useful tool for scientists working on sweeteners chemistry to understand the RS mechanism, as well as to propose and synthesize new potent sweeteners.

## MATERIALS AND METHODS

### Data Set, Data Curation, and Data Filtering

Experimental data on relative sweetness were retrieved from several published sources (refer to Table 1S for details). In order to guarantee the non-conformational QSPR model, compounds containing disconnected structures (salts) were removed due to the fact that calculations of the molecular descriptors is limited for these kind of structures. Moreover, when dealing with stereoisomers, only one of them was retained (e.g. D-rhamnose and L-rhamnose). As each compound should have a RS measured with respect to a standard solution of sucrose, a few sweeteners exhibiting a RS score measured with respect to a glucose solution were also removed. Thus, the actual data set was composed of 233 compounds. Due to the wide range of variation of relative sweetness (0.10 to 200000), the data was transformed to a logarithm value, i.e.,  $\log(RS)$  in such a way to obtain a range between -1.00 and 5.30. Finally, the SMILES (simplified molecular input line entry system) notations were obtained for each sweetener.

### Molecular Descriptors

Commonly, researchers use molecular descriptors as structural representations of molecules in order to develop QSAR/QSPR models. Descriptors are the final result of a logical and mathematical procedure that transforms chemical information encoded within a symbolic representation of a molecule into a numerical quantity or into the result of some standardized experiment (Todeschini & Consonni, 2009). The HyperChem program was used for the molecular structure representation. Subsequently, we computed 3736 non-conformational molecular descriptors using the Dragon software (version 6.0). Such descriptors were grouped into the following families: constitutional indices, functional group counts, atom-centred fragments, molecular properties, ring descriptors, topological indices, walk and path counts, connectivity indices, information indices, 2D matrix-based descriptors, 2D autocorrelations, Burden eigenvalues, P\_VSA-like descriptors, edge adjacency indices, CATS 2D, 2D atom pairs, atom-type E-state indices, and ETA indices).

### Model Development

#### *Molecular Descriptor Selection*

The selection of the most useful molecular descriptors from a large pool of correlated variables is an important goal to be addressed in the QSAR/QSPR studies. In this study, the replacement method (RM) variable subset selection (Duchowicz, Castro, & Fernández, 2006; Duchowicz, Castro, Fernández, & González, 2005) was used. This variable subset selection technique is an efficient optimization tool that generates multivariable linear regression (MLR) models by searching in a set having  $D$  descriptors for an optimal subset having  $d \ll D$  with the smallest root mean squared deviations ( $RMSD$ ). The quality of the results achieved with this technique is close to performing an exact (combinatorial) full search of molecular descriptors, although it requires much less computational work. The steps involved in the RM procedure are the following:

- A. Choose  $d$  descriptors ( $X_1, X_2, \dots, X_d$ ) randomly from a larger set of  $D$  descriptors and perform a linear regression.
- B. Select one of the descriptors of this set, say  $X_i$ , and replace it by each of the  $D$  descriptors of the pool keeping the best resulting set.
- C. Since one can start replacing any of the  $d$  descriptors in the initial model, then a regression equation with  $d$  descriptors has  $d$  possible paths to achieve the final result; e.g., the choice above will develop into path  $i$ .
- D. Choose the descriptor having the greatest relative error in its coefficient (except the one replaced in the previous step) and replace it with all of them keeping again the best set.

- E. Replace all the remaining descriptors in the same way, except for those replaced in previous steps.
- F. When finished with these steps, start again with the descriptor having the greatest relative error in the coefficient and repeat the whole process. Repeat this process as many times as necessary until the set of descriptors remains unchanged. At the end, we have the best model for the path  $i$ .
- G. Proceed in exactly the same way for all possible paths  $i = 1, 2, \dots, d$ .
- H. Finally, compare the resulting models, and retain the best one as the QSPR model for further analysis.

In addition to the use of RM for optimization, a fundamental step in the development of a QSPR model is the determination of the optimal size of the model (i.e., the number of descriptors to be included in the QSPR equation). In order to avoid overfitting in the models, an external validation set of compounds was used. In this way, we considered the quality parameter models given by the RM which are:  $R^2_{train}$ ,  $R^2_{test}$ ,  $RMSD_{train}$ ,  $RMSD_{test}$ , and  $R^2_{ij\max}$ . The best model combines the highest values for the squared correlation coefficients ( $R^2$ ), and the lowest values for both the root mean squared deviations ( $RMSD$ ) and the maximum correlation coefficient among descriptors ( $R^2_{ij\max}$ ).

### Model Validation

In order to thoroughly validate the established QSPR model, the 233 sweeteners were split into a training set (163 molecules), and an external test set (70 molecules) following a procedure based in the k-means cluster analysis (k-MCA) method (a; Rojas, Duchowicz, Tripaldi, & Pis Diez, 2015b). The training set was used to calibrate the model and to obtain its parameters and the test set was used to validate the model (i.e., to determinate its predictive ability). It is also known that the splitting procedure should be done in order to achieve similar structure-property relationships in the sets; in other words, the training set molecules should be representatives of the validation and test set compounds (Martin et al., 2012). The essence of k-MCA is to create k-clusters or groups of compounds, in such a way that compounds in the same cluster are very similar in terms of distance metrics (e.g. Euclidean distance), and compounds in different clusters are very distinct. This procedure involves the following steps:

- A. Prepare a matrix ( $\mathbf{C}$ ) that includes the log(RS) and the 2062 geometry independent molecular descriptors. This is done to account for the structure-property relationship during the clustering process. Thus, only non-conformational descriptors are used in order to avoid optimization biases. Thus, the size of  $\mathbf{C}$  is  $233 \times 2062$ .
- B. Remove the linearly dependent variables from the previous matrix. The actual size of  $\mathbf{C}$  is  $233 \times 221$ .
- C. Standardize  $\mathbf{C}$  for centering and scaling its matrix elements. This is done in order to better discern the matrix elements.
- D. Create  $N^0_{train}$  clusters with the 233 compounds through the k-MCA method, for which the  $\mathbf{C}$  matrix is used together with the Euclidean distance, and 500 runs for the numerical optimization algorithm of k-MCA in order to achieve the best solution. This step computes  $N^0_{train}$  cluster centroid locations, each centroid of  $1 \times 221$  size.  $N^0_{train} = N_{train} - N_{\min\max}$ , where  $N_{train}$  is the number of compounds to be considered in the training set and  $N_{\min\max}$  is the number of compounds that have minimum or maximum values for the experimental property.
- E. The training set is designed by including one compound per cluster, which is the compound that is nearer to the centroid in each cluster. It also includes the  $N_{\min\max}$  compounds.

- F. Create the test set by including the remaining  $N_{test} = 233 - N_{train}$  compounds through the k-MCA method in the same numerical conditions as described above.

In addition, the cross-validation technique of leave-one-out (loo) and leave-many-out (ln%o, with n% being the percentage of molecules removed from the training set) has been also performed. The statistical parameters  $R_{ln\%o}$  and  $S_{ln\%o}$  (correlation coefficient and standard deviation of leave-many-out) measure the stability of the QSPR upon inclusion/exclusion of compounds after analyzing 50000 cases for random data removal. Additionally, the Y-randomization procedure (Rücker, Rücker, & Meringer, 2007) was also applied in order to verify the model robustness. This technique is based on scrambling the experimental property values in such a way that they do not correspond to the respective compounds. At every randomization loop the model is recalculated using the selected molecular descriptors, and after analyzing 10000 cases of Y-randomization, the standard deviation ( $S^{rand}$ ) has to be a poorer value with respect to the standard deviation of the model ( $S_{train}$ ).

### *Applicability Domain Analysis*

The value of the QSPR models is related to the reliability of their predictions. The models are restricted inside a chemical space defined by chemical structures considered in the training set. Therefore, their applicability to the compounds considered in the test set is confined to those chemicals that are structurally similar to compounds present in the set used to build the model. Consequently, the applicability domain (AD) of the QSPR model was explored. The AD is a theoretically defined space that depends on the descriptors and the experimental property of the training set (Gramatica, 2007). Then, only the molecules falling within this AD are not considered model extrapolations. One possible way to analyze the AD is based on the leverage approach (Eriksson et al., 2003), which allows one to verify whether a compound never considered during the calibration can be considered as a model's interpolation (i.e., a reliable prediction), otherwise the compounds can be considered as an unreliable prediction. For this purpose, the leverage value for each compound  $i$  of the test set ( $h_i$ ) is calculated and compared with respect to a warning leverage value ( $h^*$ ) of the training set. When  $h_i > h^*$  for a given compound, it means that such compounds fall outside of the domain as a result of a substantial extrapolation, and then could not be treated as reliable. These parameters are defined in Table 2S.

### *Degree of Contribution of Selected Descriptors*

The regression coefficients ( $b_j^s$ , refer to Table 2S) were standardized in order to find out the relative contribution of the  $j$ th descriptor in the QSPR model. The larger the absolute value of  $b_j^s$ , the greater the importance of such a descriptor (Draper & Smith, 1981).

## **SOFTWARE**

HyperChem was used for molecular design. Open Babel was used to obtain the simplified molecular input line entry system (SMILES notations). Molecular descriptors were calculated by means of DRAGON version 6. Partition of the data set by means of k-MCA, variable selection by means of RM, model fitting and validation were carried out in MatLab, by using tool boxes and functions written by the authors.

## **RESULTS AND DISCUSSION**

As a first step, we calculated 3736 molecular descriptors; however, only 2062 molecular descriptors were retained by excluding descriptors with constant (1525) and near-constant (160) values, and at

least one missing value (16). Subsequently, we applied the k-MCA clustering-based procedure to split the data set of 233 sweeteners into  $N_{train} = 163$  and  $N_{test} = 70$  set compounds (refer to Table 1S), in such a way as to ensure a design with balanced sets of compounds. The  $N_{train}$  cluster centroid locations, in terms of descriptor values that minimize the squared sum of Euclidean distances of compounds located within them, are provided in a matrix labeled as C1.txt file in the Supplementary Material (<https://doi.org/10.4018/IJQSPR.2016010104>). Subsequently, the RM variable subset selection method was applied in order to build MLR models with the descriptors most correlated with the log(RS). Table 1 summarizes the best MLR models containing from 1 to 10 non-conformational descriptors where the best model is placed in bold.

For the selection of the optimal QSPR model, we analyzed simultaneously all quality parameter models given by the RM which are:  $R_{train}^2$ ,  $R_{test}^2$ ,  $RMSD_{train}$ ,  $RMSD_{test}$ ,  $R_{ij\max}^2$ , and  $d$ . The best model satisfied the highest values for the squared correlation coefficient ( $R^2$ ), and the lowest values for both the root mean squared deviation ( $RMSD$ ) and the maximum correlation coefficient among descriptors ( $R_{ij\max}^2$ ). In addition, the number of descriptors  $d$  was kept as small as possible according to the principle of parsimony (Ockham's razor) (Hoffmann, Minkin, & Carpenter, 1996), in order to avoid any possible unwarranted increased correlation between descriptors. Thus, we retained the following six-descriptor non-conformational model for further analysis:

$$\log(RS) = -2.216(\pm 0.407) + 2.002(\pm 0.246)AAC + 0.129(\pm 0.010)ATSC6p - 0.912(\pm 0.169)CATS2D\_02\_PN - 0.119(\pm 0.014)CATS2D\_05\_LL + 1.136(\pm 0.123)B07[C-N] + 0.405(\pm 0.023)ALOGP \quad (1)$$

$$N_{train} = 163, d = 6, R_{train}^2 = 0.797, S_{train} = 0.565, F = 101.915 \text{ (p-value=2.15e-51)}, \\ R_{ij\max}^2 = 0.788 \\ o(3S) = 2, R_{loo}^2 = 0.781, S_{loo} = 0.586, R_{l20\%}^2 = 0.749, S_{l20\%} = 0.629, S^{rand} = 1.126 \\ N_{test} = 70, R_{test}^2 = 0.700, S_{test} = 0.737$$

Table 1. The best QSPR models obtained by non-conformational descriptor blocks. The chosen result is placed in bold

$d$	$R_{train}^2$	$RMSD_{train}$	$R_{test}^2$	$RMSD_{test}$	$R_{ij\max}^2$	molecular descriptors
1	0.459	0.908	0.447	0.976	0.000	<i>MATSIs</i>
2	0.590	0.793	0.502	0.921	0.098	<i>SM5_B(m)</i> , <i>MATSIs</i>
3	0.678	0.705	0.633	0.796	0.221	<i>MATSIs</i> , <i>P_VSA_v_3</i> , <i>B07[C-N]</i>
4	0.726	0.652	0.623	0.816	0.989	<i>Wi_B(m)</i> , <i>VE3_B(m)</i> , <i>CATS2D_03_DD</i> , <i>B07[C-N]</i>
5	0.762	0.609	0.540	0.917	0.276	<i>GATSIs</i> , <i>P_VSA_v_3</i> <i>SssO</i> <i>CATS2D_07_AA</i> <i>F04[C-N]</i>
6	<b>0.797</b>	<b>0.565</b>	<b>0.700</b>	<b>0.737</b>	<b>0.788</b>	<i>AAC</i> , <i>ATSC6p</i> , <i>CATS2D_02_PN</i> , <i>CATS2D_05_LL</i> , <i>B07[C-N]</i> , <i>ALOGP</i>
7	0.815	0.540	0.696	0.746	0.788	<i>H%</i> , <i>ATSC6p</i> , <i>nCp</i> , <i>CATS2D_07_AA</i> , <i>CATS2D_05_LL</i> , <i>B07[C-N]</i> , <i>ALOGP</i>
8	0.822	0.533	0.699	0.751	0.946	<i>Mv</i> , <i>X3sol</i> , <i>ATSC6p</i> , <i>nCp</i> , <i>CATS2D_02_PN</i> , <i>CATS2D_05_LL</i> , <i>B07[C-N]</i> , <i>ALOGP</i>
9	0.837	0.511	0.670	0.790	0.788	<i>Mv</i> , <i>ATSC6p</i> , <i>nCp</i> , <i>C-019</i> , <i>CATS2D_07_AA</i> , <i>CATS2D_05_LL</i> , <i>B07[C-N]</i> , <i>F05[N-O]</i> , <i>ALOGP</i>
10	0.843	0.504	0.641	0.831	0.993	<i>Mp</i> , <i>VE3_H2</i> , <i>VE3_B(m)</i> , <i>nCp</i> , <i>N-074</i> , <i>CATS2D_04_AP</i> , <i>CATS2D_06_AP</i> , <i>CATS2D_06_LL</i> , <i>Hy</i> , <i>Hypnotic-80</i>



$F$  is the Fisher parameter. Meanwhile  $R_{ij\max}^2$  denotes the maximum squared correlation coefficient between pairs of descriptors. Moreover,  $o(3S)$  indicates the number of outlier compounds having a residual (difference between experimental and calculated property) greater than three-times  $S_{train}$ .

This QSPR model was predictive using the external test sets: the percentages of explained variances were  $R_{train}^2 = 79.7\%$ , and  $R_{test}^2 = 70.0\%$ . Additionally, the root mean square deviations were:  $RMSD_{train} = 0.565$ , and  $RMSD_{test} = 0.737$ . The established QSPR was also analyzed through the internal cross-validation by excluding one molecule at a time and also by removing 20% of them (33 molecules). The Y-randomization procedure illustrates that  $S_{train}(0.565) < S^{rand}(1.126)$ . Moreover, a more rigorously predictive validation of Equation 1 was performed by evaluating the predictive capability criteria suggested by Golbraikh and Tropsha (Golbraikh & Tropsha, 2002) (these parameters are also defined in Table 2S):

$$R_{loo}^2 > 0.5 \quad (0.781)$$

$$R_{test}^2 > 0.6 \quad (0.700)$$

$$1 - R_0^2 / R_{test}^2 < 0.1 \quad (0.0001)$$

$$0.85 \leq k(1.028) \leq 1.15 \text{ and } 0.85 \leq k'(0.874) \leq 1.15$$

$$R_m^2 > 0.5 \quad (0.693)$$

Thus a valid quantitative structure-property relationship was reached. Figure 1 shows the predicted log(RS) as a function of the experimental values for the training and test sets (numerical data is supplied in Table 1S). Figure 2 presents the dispersion plot of the residuals for the training and the test set, indicating that the points follow a random pattern around the zero line. These two Figures reveal that a MLR is fulfilled; i.e., there exists a tendency for the points to have a straight line trend. There are two compounds from Equation 1 that exhibit residues greater than the established limit value of  $3S (\pm 1.696)$ , which are labeled as Compound LXXX (-1.958) and N-(L-aspartyl)-1,1-diaminoalkane 5 (2.350).

Among the descriptors selected by the RM procedure, there were two descriptors for the CATS 2D block, and one descriptor for the following blocks: Information indices, Molecular Properties,

Table 2. Brief description for the non-conformational Dragon descriptors involved in the QSPR model

Name	Description	Block	Sub-Block
$AAC$	Mean information index on atomic composition	Information indices	Basic descriptors
$CATS2D\_02\_PN$	CATS2D Positive-Negative at lag 02	CATS 2D	
$CATS2D\_05\_LL$	CATS2D Lipophilic-Lipophilic at lag 05		
$ALOGP$	Ghose-Crippen octanol-water partition coefficient (logP)	Molecular properties	
$ATSC6p$	Centred Broto-Moreau autocorrelation of lag 6 weighted by polarizability	2D autocorrelations	Centred Broto-Moreau autocorrelations
$B07[C - N]$	Presence/absence of C - N at topological distance 7	2D Atom Pairs	Binary Atom Pairs of order 7



Figure 1. Experimental versus predicted log (RS) according to the established QSPR model for the 233 sweeteners

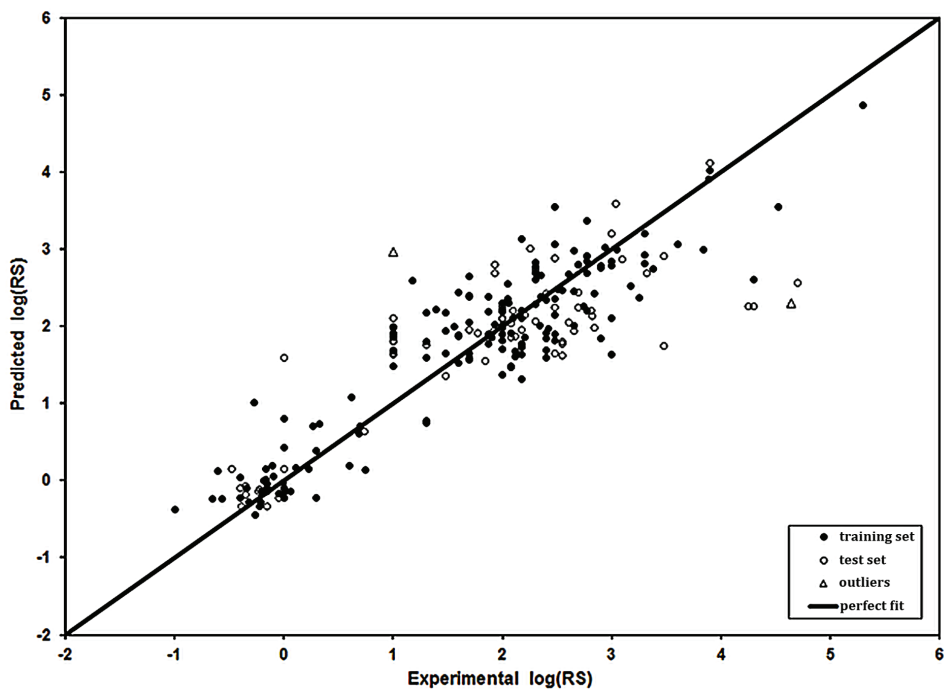
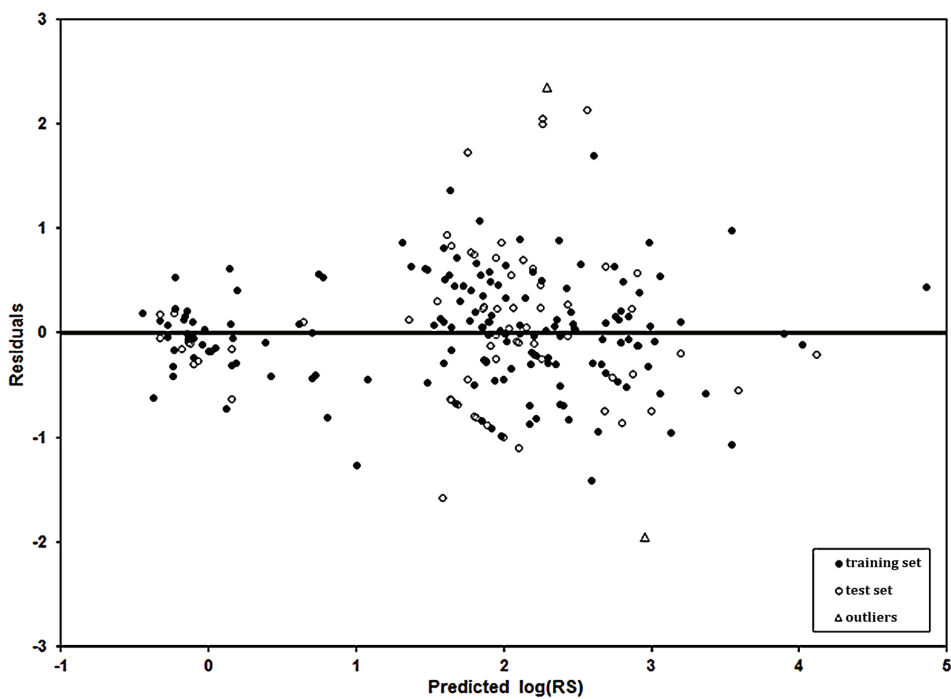


Figure 2. Dispersion plot of residuals for the QSPR model



2D autocorrelations and 2D Atom Pairs. A detail of each descriptor involved in Equation 1 is given in Table 2. These descriptors retained in the QSPR model were sufficient to study the log(RS) of these sweeteners. The moderate squared correlation coefficient between the *ATSC6p* and *CATS2D\_05\_LL* pair descriptors ( $R_{ij\max}^2 = 0.788$ ) indicated that each descriptor included different aspects of the molecular structure that succeeded in combining with the remaining variables of Equation 1 (Duchowicz, Marrugo., Vivas-Reyes, & Castro, 2010). The correlation matrix and the numerical values taken by these descriptors are included in Table 3S and Table 4S, respectively.

The relative degree of contribution of each descriptor ( $b_j^*$ ) indicates the importance or contribution of each descriptor in predicting the relative sweetness property: *ATSC6p* (1.264) > *CATS2D\_05\_LL* (0.881) > *ALOGP* (0.858) > *B07[C - N]* (0.339) > *AAC* (0.326) > *CATS2D\_02\_PN* (0.201). The data showed that four descriptors took positive numerical values of the coefficients, indicating that such variables have a synergistic effect on the RS, i.e., the larger the values of these six descriptors, the larger value of the log(RS) for a given compound. On the other hand, there were two descriptors that had a negative value of the coefficient, suggesting that this descriptor had an antagonistic influence over the prediction of this property, i.e., higher values of this descriptor for a given compound would be reduced to a lower value for such property.

The calculation of 2D autocorrelation descriptors was done by using different molecular properties in order to describe atoms at a defined lag in a molecule by considering an H-filled structure. One of these properties was the carbon-scaled atomic polarizability (*p*), which was used to weight the molecular graph in order to calculate the Centred Broto-Moreau autocorrelation at a lag 6 (*ATSC6p*) in which the centering was carried out by subtracting the value of the average property in the molecule. This descriptor describes how this property was distributed throughout the topological structure of the molecule (Todeschini & Consonni, 2009).

An Atom Pair descriptor is a simple kind of a molecular substructure defined in terms of binary data, i.e., the presence (1) or absence (0) of an established atom pair at a given topological distance (Carhart, Smith, & Venkataraghavan, 1985). Thus, the RS is related to the presence of Carbon-Nitrogen atom pair in the molecule at a topological distance of 7 (*B07[C - N]*). On the other hand, *CATS2D* descriptors were similar to the 2D Atom Pairs descriptors where the main difference was the assignment of atoms to a defined pharmacophore point type (Fechner, Franke, Renner, Schneider, & Schneider, 2003; Schneider, Neidhart, Giller, & Schmid, 1999). The lipophilic (L), positively charged or ionizable (P) and the negatively charged or ionizable (N) were used to generate the LL and PN pairs of the Potential Pharmacophore Point (PPP) pairs. Thus, *CATS2D\_05\_LL* and *CATS2D\_02\_PN* descriptors were associated with the number of intervening bonds between such PPP pairs at lag 2 and 4, respectively, as the shortest path length.

Octanol-water partition coefficient (*logP*) is a common property in computational chemistry and QSAR/QSPR modelling; a well-known model dealing with *logP* was proposed by Ghose-Crippen (Ghose, Pritchett, & Crippen, 1988; Ghose, Viswanadhan, & Wendoloski, 1998; Viswanadhan, Ghose, Revankar, & Robins, 1989). This model estimates the *logP* (*ALOGP*) as the sum of the product between the number of a particular atom type present in the molecule and the contribution of such atom type. *ALOGP* is intrinsically atomistic and easier to compute. This atomization concept of the hydrophobicity (also known as lipophilicity) is very useful in drug design.

The presence of hydrophobicity in a sweetener permits a favorable partition of the substance between the aqueous saliva fluid and the lipidic taste receptor membrane. The hydrophilicity of a sweetener allows its diffusion through the saliva to rapidly interact with the taste receptor. Therefore, the hydrophobic factor governs the sweetness intensity of molecules exhibiting an appropriate glucophore (Birch, 1987). In Equation 1 *ALOGP* indicates that the hydrophobicity of a molecule is directly related to the relative sweetness. The relationship between such descriptor and the relative sweetness was previously reported (Barker et al., 2002; Vepuri, Tawari, & Degani, 2007). In addition,

uses of  $\log P$  for describing the sweetness was reported elsewhere (Belitz, Rohse, Stempfl, & Gries, 1990; Briciu, Kot-Wasik, Wasik, Namieśnik, & Sârbu, 2010; Drew et al., 1998; Greenberg, 1980; Iwamura, 1980, 1981; Jäger, Schmidt, Schilling, & Brickmann, 2000; Spillane, Coyle, Feeney, & Thompson, 2009).

$AAC$  is the mean value of the total information content calculated from an H-filled molecular formula of a compound. Its calculation involved the total number of atoms, the number of atoms of a certain type and the probability to select in a randomly way such atom type (Todeschini & Consonni, 2009).

The model developed in the current work has a low number of uncorrelated descriptors, and could be used for explaining the RS phenomena, as well as, for predicting the RS of new sweeteners for which the leverage value should fall below the warning leverage value ( $h^* = 0.129$ ) established in the present model. The applicability domain analysis of the QSPR model shows that the compounds Periandrin III ( $h = 0.153$ ) and Aspartic acid fenchyl ester ( $h = 0.133$ ) fall outside this theoretical domain.

According to Figure 1, there are some compounds that show a high deviation from the calculated  $\log(RS)$  with respect to the experimental values obtained by tasting panels, especially at larger values of this property. We believe that these particular deviations may be ascribed to the difficulty to model this sensory property due to the fact that there are intrinsic human errors during RS determinations. These differences may occur through the use of untrained or undertrained people assigned to tasting panels, or by assigning too few people to these panels, as well as, to the fact that humans are unlikely to discern differences of sweetness due to receptor pre-saturation on the taste buds of the tongue (G. Birch & Mylvaganam, 1976). On the other hand, there is a wide heterogeneity of molecules considered in this data set which were synthesized and measured by several research groups. In fact, some QSPR studies regarding RS responses are usually carried out by considering only data sets of homogeneous families or a few families of compounds, limiting their ability to generalize these models to other kinds of sweeteners.

A comparison of published models for the prediction of the RS is presented in Table 3. Model of Equation 1 has reduced number of descriptors ( $d$ ) and similar performances in calibration and prediction with respect to the reported MLR models, as well as to the artificial neural network (ANN) (Yang et al., 2011) and support Vector Machine (SVM) (Zhong et al., 2013) non-linear models. Moreover, its predictive feature demonstrates to be higher than the 3D-QSAR models using Genetic Functional Algorithm (GFA), Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) (Vepuri et al., 2007), which exhibit the highest  $RMSE$ . On the other hand, external validation of proposed QSPR models was not performed in the some of the published studies (Katritzky, Lobanov, Karelson, 1995; Kelly, Spillane, & Newell, 2005; Tarko, Lupescu, & Constantinescu-Groposila, 2006) and therefore a comparison of predictive capabilities with respect to the model proposed in this study is not feasible.

## CONCLUSION

In this study, a conformation-independent QSPR model was performed to predict the relative sweetness of 233 sweeteners. The use of the RM variable subset selection technique allowed the selection of the optimal subset of six Dragon descriptors. In fact, the mathematical model demonstrates good performance in fitting and cross-validation, as well as accurate prediction of the test set of molecules. In addition, the similar performances for the training set, cross-validation and test set indicates the absence of overfitting. It is interesting to highlight that Dragon molecular descriptors and the RM approach were used for the first time to modelling the RS property. This model could be useful for scientists to rationally design new potent sweeteners. Finally, the non-conformational QSPR

**Table 3. Comparison among the largest and recently published QSAR/QSPR models for the prediction of the relative sweetness**

Reference	Model	Number of Compounds	d	$R^2_{train}$	$RMSD_{train}$	$R^2_{test}$	$RMSD_{test}$
(Katritzky et al., 2002)	MLR	151	5	0.842	0.108	-- <sup>a</sup>	--
		238	5	0.686	0.098	--	--
(Kelly et al., 2005)	CART	83	6	0.627	--	--	--
(Tarko, Lupescu, & Groposila-Constantinescu, 2005)	MLR	123	4	--	0.485	--	0.507
(Tarko et al., 2006)	MLR	136	4	0.655	0.652	--	--
		121	5	0.847	0.399	--	--
(Vepuri et al., 2007)	MLR (3D-GFA)	53	4	0.753	5.148	0.375	--
	MLR (3D-CoMFA)		3	0.768		0.535	--
	MLR (3D-CoMSIA)		6	0.927		0.596	--
(Yang et al., 2011)	MLR	103	3	0.885	1.023	0.856	1.205
	ANN			0.899	1.049	0.869	1.162
	SVM			0.910	1.037	0.889	1.192
(Zhong et al., 2013)	MLR	320	12	0.814	0.958	0.773	1.029
	SVM			0.830	0.979	0.778	0.994
Rojas, Tripaldi, & Duchowicz	MLR	233	6	0.797	0.565	0.700	0.737

<sup>a</sup>not available

methodology continues to emerge as an efficient alternative approach to develop models based on topological and constitutional molecular aspects of chemical compounds.

## ACKNOWLEDGMENT

Cristian Rojas is grateful for his PhD Fellowship from the National Secretary of Higher Education, Science, Technology and Innovation (SENESCYT) from the Republic of Ecuador. Pablo R. Duchowicz wishes to thank the National Scientific and Technical Research Council of Argentina (CONICET) for the project grant PIP11220100100151, and the Minister of Science, Technology and Productive Innovation for the use of the electronic library facilities. Pablo R. Duchowicz is a member of the Scientific Researcher Career of CONICET.

## REFERENCES

- Arnoldi, A., Bassoli, A., Borgonovo, G., Drew, M. G., Merlini, L., & Morini, G. (1998). Sweet isovanillyl derivatives: Synthesis and structure-taste relationships of conformationally restricted analogues. *Journal of Agricultural and Food Chemistry*, 46(10), 4002–4010. doi:10.1021/jf980246b
- Arnoldi, A., Bassoli, A., Borgonovo, G., & Merlini, L. (1995). Synthesis and sweet taste of optically active (-)-haematoxylin and of some ( $\pm$ )-haematoxylin derivatives. *Journal of the Chemical Society, Perkin Transactions 1: Organic and Bio-Organic Chemistry*, (19): 2447–2453. doi:10.1039/p19950002447
- Arnoldi, A., Bassoli, A., Borgonovo, G., Merlini, L., & Morini, G. (1997). Synthesis and structure-activity relationships of sweet 2-benzoylbenzoic acid derivatives. *Journal of Agricultural and Food Chemistry*, 45(6), 2047–2054. doi:10.1021/jf960755i
- Arnoldi, A., Bassoli, A., & Merlini, L. (1996). Progress in isovanillyl sweet compounds. *Food Chemistry*, 56(3), 247–253. doi:10.1016/0308-8146(96)00021-0
- Arnoldi, A., Bassoli, A., Merlini, L., & Ragg, E. (1991). Isovanyll sweeteners. Synthesis, conformational analysis, and structure-activity relationship of some sweet oxygen heterocycles. *Journal of the Chemical Society, Perkin Transactions 2: Physical Organic Chemistry*, (9): 1399–1406. doi:10.1039/p29910001399
- Arnoldi, A., Bassoli, A., Merlini, L., & Ragg, E. (1993). Isovanyll sweeteners. Synthesis and sweet taste of sulfur heterocycles. *Journal of the Chemical Society, Perkin Transactions 1: Organic and Bio-Organic Chemistry*, (12): 1359–1366. doi:10.1039/p19930001359
- Barker, J. S., Hattotuwigama, C. K., & Drew, M. G. B. (2002). Computational studies of sweet-tasting molecules. *Pure and Applied Chemistry*, 74(7), 1207–1217. doi:10.1351/pac200274071207
- Bassoli, A., Borgonovo, G., Drew, M. G., & Merlini, L. (2000). Enantiodifferentiation in taste perception of isovanillic derivatives. *Tetrahedron, Asymmetry*, 11(15), 3177–3186. doi:10.1016/S0957-4166(00)00271-8
- Bassoli, A., Drew, M. G. B., Hattotuwigama, C. K., Merlini, L., Morini, G., & Wilden, G. R. H. (2001). Quantitative structure-activity relationships of sweet isovanillyl derivatives. *Quantitative Structure-Activity Relationships*, 20(1), 3–16. doi:10.1002/1521-3838(200105)20:1<3::AID-QSAR3>3.0.CO;2-H
- Bassoli, A., Drew, M. G. B., Merlini, L., & Morini, G. (2002). General pseudoreceptor model for sweet compounds: A semiquantitative prediction of binding affinity for sweet-tasting molecules. *Journal of Medicinal Chemistry*, 45(20), 4402–4409. doi:10.1021/jm020833v PMID:12238920
- Belitz, H.-D., Grosch, W., & Schieberle, P. (2009). *Food chemistry* (4th ed.). Heidelberg: Springer-Verlag.
- Belitz, H.-D., Rohse, H., Stempfl, W., & Gries, H. (1990). Trihalogenated benzamides: Structure taste relationships. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 190(4), 319–324. doi:10.1007/BF01184500
- Birch, G., & Mylvaganam, A. (1976). Evidence for the proximity of sweet and bitter receptor sites.
- Birch, G. G. (1987). Sweetness and sweeteners. *Endeavour*, 11(1), 21–24. doi:10.1016/0160-9327(87)90165-7 PMID:2436888

- Briciu, R. D., Kot-Wasik, A., Wasik, A., Namieśnik, J., & Sârbu, C. (2010). The lipophilicity of artificial and natural sweeteners estimated by reversed-phase thin-layer chromatography and computed by various methods. *Journal of Chromatography. A*, 1217(23), 3702–3706. doi:10.1016/j.chroma.2010.03.057 PMID:20430396
- Carhart, R. E., Smith, D. H., & Venkataraghavan, R. (1985). Atom pairs as molecular features in structure-activity studies: Definition and applications. *Journal of Chemical Information and Computer Sciences*, 25(2), 64–73. doi:10.1021/ci00046a002
- deMan, J. M. (1999). *Principles of food chemistry* (3rd ed.). Springer. doi:10.1007/978-1-4614-6390-0
- Deutsch, E. W., & Hansch, C. (1966). Dependence of relative sweetness on hydrophobic bonding. *Nature*, 211(5044), 75. doi:10.1038/211075a0 PMID:5967471
- Diudea, M. V. (2001). *QSPR/QSAR studies by molecular descriptors*. New York: Nova Science Publishers.
- Dragon (version 6) (Version 6.0). (2014): software for molecular descriptor calculation, TALETE, srl., <http://www.talete.mi.it/>
- Draper, N. R., & Smith, H. (1981). *Applied regression analysis*. New York.
- Drew, M. G. B., Wilden, G. R. H., Spillane, W. J., Walsh, R. M., Ryder, C. A., & Simmie, J. M. (1998). Quantitative structure-activity relationship studies of sulfamates RNHSO<sub>3</sub>Na: Distinction between sweet, sweet-bitter, and bitter molecules. *Journal of Agricultural and Food Chemistry*, 46(8), 3016–3026. doi:10.1021/jf980095c
- Duchowicz, P. R., Castro, E. A., & Fernández, F. M. (2006). Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *MATCH Commun. Math. Comput. Chem.*, 55, 179–192.
- Duchowicz, P. R., Castro, E. A., Fernández, F. M., & González, M. P. (2005). A new search algorithm of QSPR/QSAR theories: Normal boiling points of some organic molecules. *Chemical Physics Letters*, 412(4-6), 376–380. doi:10.1016/j.cplett.2005.07.016
- Duchowicz, P. R., Marrugo, H. J. J., Vivas-Reyes, R., & Castro, E. A. (2010). QSPR applied on gas chromatography indices of polycyclic aromatic compounds. *International Journal of Environmental Sciences*, 1, 73–77.
- Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives*, 111(10), 1361–1375. doi:10.1289/ehp.5758 PMID:12896860
- Fechner, U., Franke, L., Renner, S., Schneider, P., & Schneider, G. (2003). Comparison of correlation vector methods for ligand-based similarity searching. *Journal of Computer-Aided Molecular Design*, 17(10), 687–698. doi:10.1023/B:JCAM.0000017375.61558.ad PMID:15068367
- Ghose, A. K., Pritchett, A., & Crippen, G. M. (1988). Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions. *Journal of Computational Chemistry*, 9(1), 80–90. doi:10.1002/jcc.540090111
- Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1998). Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of ALOGP and CLOGP methods. *The Journal of Physical Chemistry A*, 102(21), 3762–3772. doi:10.1021/jp980230o
- Golbraikh, A., & Tropsha, A. (2002). Beware of q<sup>2</sup>! *Journal of Molecular Graphics & Modelling*, 20(4), 269–276. doi:10.1016/S1093-3263(01)00123-1 PMID:11858635
- Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, 26(5), 694–701. doi:10.1002/qsar.200610151
- Greenberg, M. J. (1980). The importance of hydrophobic properties of organic compounds on their taste intensities: A quantitative structure-taste-intensity study. *Journal of Agricultural and Food Chemistry*, 28(3), 562–566. doi:10.1021/jf60229a049
- Hansch, C. (1970). Use of  $\sigma^+$  in structure-activity correlations. *Journal of Medicinal Chemistry*, 13(5), 964–966. doi:10.1021/jm00299a037 PMID:5458389
- Hoffmann, R., Minkin, V. I., & Carpenter, B. K. (1996). Ockham's razor and chemistry. *Bulletin de la Société Chimique de France*, 133(2), 117–130.



- Iwamura, H. (1980). Structure-taste relationship of perillartine and nitro- and cyanoaniline derivatives. *Journal of Medicinal Chemistry*, 23(3), 308–312. doi:10.1021/jm00177a020 PMID:7365747
- Iwamura, H. (1981). Structure-sweetness relationship of L-aspartyl dipeptide analogs. A receptor site topology. *Journal of Medicinal Chemistry*, 24(5), 572–583. doi:10.1021/jm00137a018 PMID:7241515
- Jäger, R., Schmidt, F., Schilling, B., & Brickmann, J. (2000). Localization and quantification of hydrophobicity: The molecular free energy density (MolFESD) concept and its application to sweetness recognition. *Journal of Computer-Aided Molecular Design*, 14(7), 631–646. doi:10.1023/A:1008181611372 PMID:11008885
- Katritzky, A. R., Lobanov, V. S., & Karelson, M. (1995). QSPR: The correlation and quantitative prediction of chemical and physical properties from structure. *Chemical Society Reviews*, 24(4), 279–287. doi:10.1039/cs9952400279
- Katritzky, A. R., Petrukhin, R., Perumal, S., Karelson, M., Prakash, I., & Desai, N. (2002). A QSPR study of sweetness potency using the CODESSA program. *Croatica Chemica Acta*, 75(2), 475–502.
- Kelly, D. P., Spillane, W. J., & Newell, J. (2005). Development of structure-taste relationships for monosubstituted phenylsulfamate sweeteners using classification and regression tree (CART) analysis. *Journal of Agricultural and Food Chemistry*, 53(17), 6750–6758. doi:10.1021/jf0507137 PMID:16104795
- Martin, T. M., Harten, P., Young, D. M., Muratov, E. N., Golbraikh, A., Zhu, H., & Tropsha, A. (2012). Does rational selection of training and test sets improve the outcome of QSAR modeling? *Journal of Chemical Information and Modeling*, 52(10), 2570–2578. doi:10.1021/ci300338w PMID:23030316
- Nanayakkara, N. P. D., Hussain, R. A., Pezzuto, J. M., Soejarto, D. D., & Kinghorn, A. D. (1988). An intensely sweet dihydroflavonol derivative based on a natural product lead compound. *Journal of Medicinal Chemistry*, 31(6), 1250–1253. doi:10.1021/jm00401a030 PMID:3373494
- O'Brien-Nabors, L. (2001). *Alternative sweeteners* (3rd ed.). Marcel Dekker, Inc.
- Organisation for Economic Co-operation and Development. (2007). Guidance document on the validation of (Quantitative)Structure-Activity Relationships [(Q)SAR] models.
- Pietrzycki, W. (2001). QSAR computational model of Nofre-Tinti theory on sweetness of mono- and disaccharides composed by pyranose units. *Polish Journal of Chemistry*, 75(10), 1569–1582.
- Rao, M., & Kumar, N. (1986). Quantitative correlation between steric factors and the relative sweetness of 2-substituted 5-nitroanilines. *Indian Drugs*, 23, 35–38.
- Rojas, C., Duchowicz, P. R., Pis Diez, R., & Tripaldi, P. (2016). Applications of quantitative structure-relative sweetness relationships in food chemistry. In A. G. Mercader, P. R. Duchowicz, & P. M. Sivakumar (Eds.), *Chemometrics applications and research: QSAR in medicinal chemistry*. CRC Press, Taylor & Francis Group.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015a). QSPR analysis for the retention index of flavors and fragrances on a OV-101 column. *Chemometrics and Intelligent Laboratory Systems*, 140(0), 126–132. doi:10.1016/j.chemolab.2014.09.020
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015b). Quantitative structure–property relationship analysis for the retention index of fragrance-like compounds on a polar stationary phase. *Journal of Chromatography. A*, 1422, 277–288. doi:10.1016/j.chroma.2015.10.028
- Rücker, C., Rücker, G., & Meringer, M. (2007). Y-Randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, 47(6), 2345–2357. doi:10.1021/ci700157b PMID:17880194
- Schneider, G., Neidhart, W., Giller, T., & Schmid, G. (1999). “Scaffold hopping” by topological pharmacophore search: A contribution to virtual screening. *Angewandte Chemie International Edition*, 38(19), 2894–2896. doi:10.1002/(SICI)1521-3773(19991004)38:19<2894::AID-ANIE2894>3.0.CO;2-F PMID:10540384
- Shallenberger, R. S. (1993). *Taste chemistry*. Springer Science & Business Media. doi:10.1007/978-1-4615-2666-7
- Singh, R. K., Khan, M. A., & Singh, P. P. (2014). Rating of sweetness by molar refractivity and ionization potential: QSAR study of sucrose and guanidine derivatives. *South African Journal of Chemistry. Suid-Afrikaanse Tydskrif vir Chemie*, 67, 12–20.



Spillane, W. J. (1983). *Quantitative structure-taste relationship studies of sulphamate sweeteners* (pp. 16–19). London: Chem. and Ind.

Spillane, W. J., Coyle, C. M., Feeney, B. G., & Thompson, E. F. (2009). Development of structure-taste relationships for thiazolyl-, benzothiazolyl-, and thiadiazolylsulfamates. *Journal of Agricultural and Food Chemistry*, 57(12), 5486–5493. doi:10.1021/jf9002472 PMID:19456131

Spillane, W. J., & McGlinchey, G. (1981). Structure-activity studies on sulfamate sweeteners II: Semiquantitative structure-taste relationship for sulfamate (RNHSO<sub>3</sub>) sweeteners-the role of R. *Journal of Pharmaceutical Sciences*, 70(8), 933–935. doi:10.1002/jps.2600700826 PMID:7310667

Spillane, W. J., McGlinchey, G., Muirheartaigh, I. Ó., & Benson, G. A. (1983). Structure-activity studies on sulfamate sweeteners III: Structure-taste relationships for heterosulfamates. *Journal of Pharmaceutical Sciences*, 72(8), 852–856. doi:10.1002/jps.2600720804 PMID:6620135

Spillane, W. J., Ryder, C. A., Curran, P. J., Wall, S. N., Kelly, L. M., Feeney, B. G., & Newell, J. (2000). Development of structure-taste relationships for sweet and non-sweet heterosulfamates. *Journal of the Chemical Society, Perkin Transactions 2: Physical Organic Chemistry*, (7): 1369–1374. doi:10.1039/b0024821

Spillane, W. J., Ryder, C. A., Walsh, M. R., Curran, P. J., Concagh, D. G., & Wall, S. N. (1996). Sulfamate sweeteners. *Food Chemistry*, 56(3), 255–261. doi:10.1016/0308-8146(96)00022-2

Spillane, W. J., & Sheahan, M. B. (1989). Semi-quantitative and quantitative structure-taste relationships for carboand hetero-sulphamate (RNHSO<sub>3</sub>) sweeteners. *Journal of the Chemical Society, Perkin Transactions 2: Physical Organic Chemistry*, (7): 741–746. doi:10.1039/p29890000741

Tarko, L., Lupescu, I., & Constantinescu-Groposila, D. (2006). QSAR Studies on amino-succinamic acid derivatives sweeteners. *ARKIVOC*, 13, 22–40.

Tarko, L., Lupescu, I., & Groposila-Constantinescu, D. (2005). Sweetness power QSARs by PRECLAV software. *ARKIVOC*, 10, 254–271.

Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics* (Vol. 2). Weinheim: Wiley-VCH. doi:10.1002/9783527628766

Trinajstić, N. (1992). *Chemical graph theory*. Boca Raton, FL: CRC Press.

van der Heijden, A., Brussel, L. B. P., & Peer, H. G. (1979). Quantitative structure-activity relationships (QSAR) in sweet aspartyl dipeptide methyl esters. *Chemical Senses*, 4(2), 141–152. doi:10.1093/chemse/4.2.141

Vepuri, S. B., Tawari, N. R., & Degani, M. S. (2007). Quantitative structure-activity relationship study of some aspartic acid analogues to correlate and predict their sweetness potency. *QSAR & Combinatorial Science*, 26(2), 204–214. doi:10.1002/qsar.200530191

Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., & Robins, R. K. (1989). Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *Journal of Chemical Information and Computer Sciences*, 29(3), 163–172. doi:10.1021/ci00063a006

Walters, D. E. (2002). Genetically evolved receptor models (GERM) as a 3D QSAR tool. In H. Kubinyi, G. Folkers, & Y. C. Martin (Eds.), *3D QSAR in Drug Design. Recent Advances* (Vol. III, pp. 159-166). Kluwer Academic Publishers.

Walters, D. E. (2006). Analysing and predicting properties of sweet-tasting compounds. In W. J. Spillane (Ed.), *Optimising sweet taste in foods* (pp. 283–291). doi:10.1533/9781845691646.3.283

Walters, D. E., & Hinds, R. M. (1994). Genetically evolved receptor models: A computational approach to construction of receptor models. *Journal of Medicinal Chemistry*, 37(16), 2527–2536. doi:10.1021/jm00042a006 PMID:8057298

Yamato, M., & Hashigaki, K. (1979). Chemical structure and sweet taste of isocoumarins and related compounds. *Chemical Senses*, 4(1), 35–47. doi:10.1093/chemse/4.1.35