QSAR Analysis for the Inhibition of the Mutagenic Activity by Anthocyanin Derivatives

Nicolas Alejandro Szewczuk, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas, Argentina

Pablo Román Duchowicz, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas, Argentina

iD https://orcid.org/0000-0003-3225-7495

Alicia Beatriz Pomilio, Universidad de Buenos Aires, Argentina

ABSTRACT

Flavonoid compounds modulate the cytochrome P450 3A4 enzyme activity and inhibit the mutagenic activity of mammalian cells, preventing carcinogen activation and cellular DNA damage. In this work, the quantitative structure-activity relationships (QSAR) theory is applied to predict the cytochrome P450 3A4 inhibition constant by anthocyanin derivatives. Different freely available software calculates 102,260 non-conformational molecular descriptors. A training set of 12 compounds is used to calibrate the best univariable linear regression models, while a test set of 4 compounds is used to explore their predictive capability. The present results are compared with previously reported ones by using 3D-QSAR, thus demonstrating that the proposed topological QSAR models achieve acceptable statistical quality. The proposed model provides a prospective QSAR guide for the search of new anthocyanin derivatives possessing high or low predicted mutagenicity.

KEYWORDS

Anthocyanins, Cytochrome P450 3A4, Molecular Descriptors, Mutagen, Quantitative Structure-Activity Relationships

1. INTRODUCTION

The family of cytochrome P450 (CYP) proteins is widely distributed in all kingdoms of life, from bacteria, archaea, and viruses to higher plants and animals. The substrates of CYP proteins can be very varied, including endogenous compounds produced by cell metabolism, and exogenous substances such as biogenic amines (Sánchez-Jiménez et al., 2013), cytostatic agents (Evteev et al., 2006), and steroids (Yamazaki et al., 1998). CYPs are monooxygenases that contain a heme group, which catalize a variety of reactions, such as hydroxylation, epoxidation, oxidation, reduction, deamination, dehalogenation, dealkylation, dehydrogenation, and demethoxygenation (Ortiz de Montellano&De Voss, 2005; Gillam&Hunter, 2007).

CYP3A4 is the most abundant human isoform. CYP3A4 is found in the liver and it metabolizes, completely or partially, around 45-60% of clinical drugs (Li et al., 2018). Therefore, the regulation

of its expression and/or activity is clinically relevant (Zanger&Schwab, 2013). Some members of the flavonoid family, including proanthocyanidins (Lobayan et al., 2009) and bioflavonoids (Mercader&Pomilio, 2011), that are present in fruits, vegetables, and dietary supplements, have demonstrated to possess the ability to modify the activity of the CYP system (Miron et al., 2017), thus changing the pharmacodynamics and pharmacokinetics of many commercial drugs (Moon et al., 2006; Dreiseitel et al., 2008).

Anthocyanins belong to the family of flavonoids, being widely distributed in nature as hydrosoluble pigments responsible for red, blue, purple, and violet colors of flowers, bracts, leaves, fruits, and also stems (Cheynier, et al., 2015; Pomilio&Mercader, 2018). Anthocyanins are mainly found as glycosylated derivatives of the anthocyanidins cyanidin, delphinidin, malvidin, pelargonidin, peonidin, and petunidin (Pomilio&Mercader, 2018). The oral administration of anthocyanin-containing fruits, extracts, and pure compounds has shown to be effective in the prevention and/or treatment of some diseases, such as cardiovascular disorders, diabetes, arthritis, and cancer, due to the antioxidant and anti-inflammatory properties (Bentz et al., 2017; Putta et al., 2018; Chen et al., 2018; Pomilio&Mercader, 2018). A part of the efficiency of chemotherapeutic and antimicrobial agents after the administration of anthocyanins and derivatives is due to their positive activity in multidrug resistance (Bentz et al., 2017; Pomilio&Mercader, 2018).

Previously, Dreiseitel et al., 2008 reported that this kind of compounds and their derivatives, for example, aglycons and procyanidins have shown concentration-dependent inhibition of CYP. On the other hand, the same authors suggested that the number of sugar moieties showed a decrease in the effects of anthocyanidins on CYP3A4, which underscores the need for more detailed data on structure-activity relationships. The anthocyanins isolated from the fruits of *Aronia melanocarpa* have the ability to inhibit the mutagenic activity of α-benzopyrene and 2-aminofluorene in the Ames test (Gasiorowski et al., 1997), thus demonstrating the antimutagenic properties of these compounds. Furthermore, the ability to inhibit CYP of this group of compounds could be considered antimutagenic due to the fact that these enzymes contribute to the carcinogenesis in several organs, and increase toxicity and DNA damage.

This has been the goal of many modeling studies (Langowski&Long, 2002; Kirton, 2002), based on the framework mentioned above, and the crucial role of the CYP enzyme set. Therefore, interest has arisen in performing a predictive and selective search within the family of anthocyanins, several precursors and derivatives, through studies of Quantitative Structure-Activity Relationships (QSAR) (Bentz, et al., 2017; Duchowicz et al., 2019), to obtain a mathematical model with predictive power by relating the molecular structure, encoded within the so-called molecular descriptors (Todeschini&Consonni, 2009), to a biological activity of interest (Pomilio et al., 2019).

The present work comes from these considerations, and has as aim the development of novel, simple and easily interpretable QSAR systems to evaluate a series of anthocyanin derivatives as CYP3A4 inhibitors by analyzing and comparing the experimental values of inhibition constant (K_i) of CYP3A4 with those obtained by means of predictive models.

The Multivariable Linear Regression (MLR) analysis constitutes a widely employed mathematical technique that has been extensively used in past QSAR studies for establishing the hypothetical connection between the chemical structures of compounds to their activities of interest (Ortiz et al., 2017; Fioressi et al., 2019). In this study we establish univariable linear regression QSAR models for predicting the P450 3A4 inhibitory activity in a molecular set composed of 16 anthocyanins shown in Figure 1. The so established models are afterwards compared with reported 3D QSAR models.

Figure 1. Molecular structures of 16 anthocyanin derivatives with antimutagenic activity*
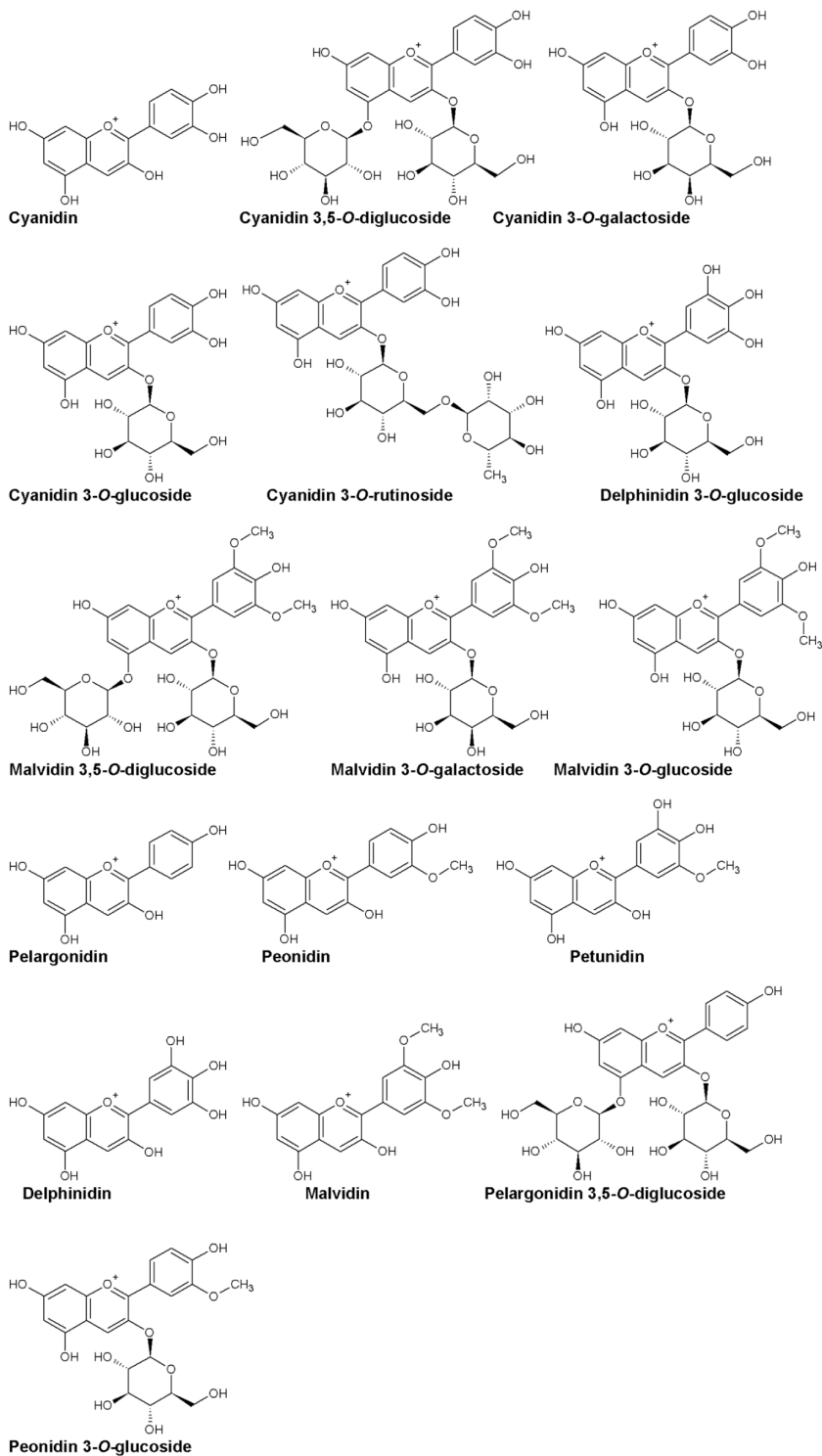
* Conformational features of sugars, such as chairs, have not been included according to QSAR studies.

2. METHODS AND MATERIALS

2.1. Experimental Data

The experimental P450 3A4 inhibitory activities are extracted from the literature (Shityakov et al., 2014) and expressed as the negative decimal logarithm of the experimental inhibition constant (pK_i), which has been obtained by using the pIC_{50} value calculator (Selvaraj et al., 2011) according to the Michaelis-Menten enzyme-substrate interaction equation for noncompetitive species. The pK_i parameter ranges in the interval (3.6030, 4.9100). The chemical names of the studied compounds together with their experimental data are provided in Table 3.

2.2. Structural Representation and Molecular Descriptors Calculation

The chemical structures of the 16 anthocyanin derivatives shown in Figure 1 are drawn with the ACDLabs/ChemSketch freeware (2019) with molecules in MDL mol (V2000) format. All file format conversions are performed with Open Babel for Windows (2019). The conformation-independent conventional molecular descriptors are computed as follows.

We use the Pharmaceutical Data Exploration Laboratory (PaDEL) freeware version 2.20 (2019), because it has the advantage that it is a freely available and open source program. PaDEL allows calculating 17,536 0D-2D molecular descriptors and fingerprint types.

The PyDescriptor freeware (Masand&Rastija, 2017) can calculate a diverse pool of 16,299 molecular descriptors comprising easily understandable 1D-2D descriptors encoding pharmacophoric patterns, atomic fragments and a variety of fingerprints.

Other molecular descriptors are calculated with the Molecular Descriptors from 2D structures (Mold^2) freeware (Hong et al., 2008), which generates 777 1D-2D structural variables with molecules in MDL sdf format.

More 2D molecular descriptors are calculated with the Quadratic, Bilinear and N-Linear MapS (QuBiLs-MAS) suite (Valdés-Martiní et al., 2017) by using the Graph-Theoretic Electronic-Density Matrices and Atomic Weightings (MAS) module from the ToMoCoMD-CARDD free multi-platform freeware. The QuBiLs-MAS algebraic module calculates 67,584 molecular descriptors as Quadratic, Bilinear and Linear Maps based on Pseudograph-Theoretic Electronic-Density Matrices and Atomic Weightings. The calculated invariants are: 'euclidean distance', 'arithmetic mean', and 'standard deviation'.

Finally, the ISIDA/Fragmentor freeware (2017) is used to count atoms and linear structural fragments of 2 to 5 atoms length, thus generating 64 possible atom and fragment types in the present data set, with molecules in MDL sdf format.

Therefore, the total number of non-conformational molecular descriptors explored in this work is 102,260, calculated on the cationic anthocyanin derivatives, and capturing the most relevant structural characteristics affecting the studied antimutagenic activity.

2.3. Model Search

First of all, the set of 102,260 non-conformational molecular descriptors is analyzed in order to remove the 'collinear' descriptors, having duplicated structural information. Thus, linearly dependent descriptor pairs are identified, and only one variable is kept from each pair for further analysis, reducing the pool size to 47,117 linearly independent descriptors.

The empirical 'rule of thumb' (Tute, 1971), widely used in many QSAR studies when few experimental data are available, indicates that more than six experimental observations per variable should be used during the model development. With the purpose of avoiding overfitting problems and chance correlation, we decide to establish univariable linear regression models. The best descriptors are searched that lead to the smallest value for the root mean square error (*RMS*) or the model's standard deviation (*S*) in the training set.

The 16 anthocyanin molecules are partitioned into a training set (12 compounds, 75%) for model fitting and a test set (4 compounds, 25%) for model validation. For comparison purposes, this is done in two alternative ways: a) using the same methodology as in the previous study of Shityakov and collaborators (Shityakov et al., 2014), that is to say, using a random selection method that is part of the Strike 1.9 module integrated in the Maestro 9.1 molecular modeling suite (2019); and b) by means of the Balanced Subsets Method (BSM) (Rojas et al., 2015). The BSM is a procedure developed by our group that ensures that balanced subsets are generated, in such a way that training and test sets are not defined with random molecules but involve similar structure-property relationships for each set.

Apart from using a test set, each model is theoretically validated by means of a 'Leave-One-Out' (loo) cross-validation. Further validation parameters are calculated as proposed in the literature (Chirico and Gramatica, 2012; Roy et al., 2013).

All the MATLAB R2013a programmed algorithms used in our calculations have been developed in our research group and are available upon request.

3. RESULTS AND DISCUSSION

From the pool of 47,117 linearly independent descriptors, different single variable linear regression models are first identified, by searching for the smallest values of the root mean square (RMS) error parameter in the training set. After this initial optimization step, then different model evaluation criteria are considered for selecting the best linear model, such as the RMS value in the test set, cross-validation results, number of outlier compounds, and other theoretical validation strategies.

The random partition of the dataset as used in (Shityakov et al., 2014) leads to the following one descriptor QSAR model, where the best non-conformational structural variable is searched among 47,117 available molecular descriptors:

$$pK_i = -0.0029\, d_1 + 5.33 \tag{1}$$
$$N_{train} = 12\,,\ R^2_{train} = 0.97\,,\ RMS_{train} = 0.07$$
$$R^2_{loo} = 0.95\,,\ RMS_{loo} = 0.09\,,\ o2.5 = 0\,,\ R^2_{rand} = 0.94\,,\ RMS^{rand} = 0.09\,,\ R^2_m = 0.86$$
$$N_{test} = 4\,,\ R^2_{test} = 0.98\,,\ RMS_{test} = 0.10$$

Equation 1 does not involve training set compounds with very high residuals because the number of outlier compounds in such set with residual (difference between experimental and predicted activity) greater than 2.5 times S_{train} ($o2.5$) is zero. A plot for the pK_i predictions of Equation 1 and the experimental values is provided in Figure 2A. Table 3 includes the predictions for the 16 studied anthocyanin derivatives and indicates the molecules composing the test set.

The Y-Randomization technique (Rücker et al., 2007) demonstrates through 10,000 cases of experimental activity scrambling that the model has $RMS_{train} < RMS^{rand}$ and $R^2_{rand} < R^2_{train}$, and consequently that a valid structure-pK_i relationship is established that does not involve chance correlation. The reported R^2_{rand} and RMS^{rand} values are for the worst possible case after analyzing 10,000 cases. In addition, the internal validation process of the leave-one-out (loo) cross validation technique (Golbraikh&Tropsha, 2002) succeeds in predicting one molecule excluded at a time from the training set ($R^2_{loo} > 0.5$). The R^2_m index (Roy et al., 2009) is also acceptable for Equation 1 as $R^2_m > 0.5$.

The QSAR of Equation 1 has an acceptable predictive power on the external test set of 4 anthocyanins, and clearly demonstrates its ability to predict the inhibition constant on compounds not considered during the model calibration. It is found that the 4 test set compounds belong to the

Figure 2. Predicted and experimental inhibition constants for anthocyanin derivatives according to: (A) Equation 1; and (B) Equation 2
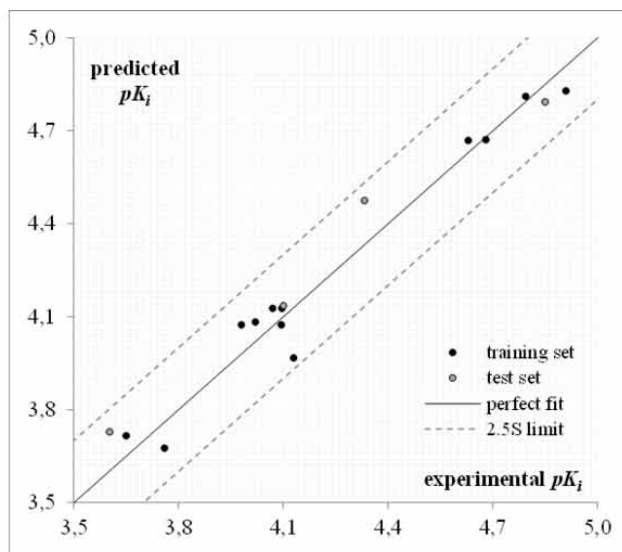

A)
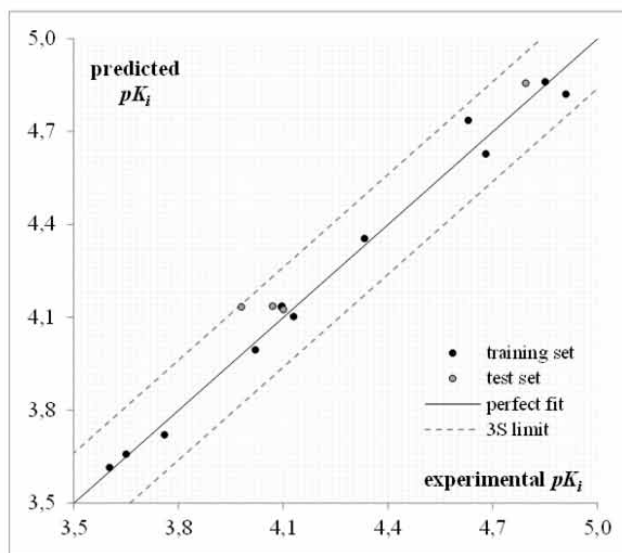
B)

model's applicability domain (AD) (Eriksson et al., 2003), as their leverage (h_i) values fall under the warning leverage limit (h^*) of 0.5. The Williams plot for Equation 1 (standardized residuals as function of the h_i values) is provided in Figure 3A.

The d_1 molecular descriptor is calculated with the QuBiLs-MAS freeware, being denoted with the following symbol within this program: N2_Q_AB_nCi_2_DS14_T_KA_psa_MAS. Such notation means that it is the 14th order double stochastic atom-based non-chiral total (global) quadratic index,

Figure 3. Standardized residuals as function of the leverage value for: (A) Equation 1; and (B) Equation 2
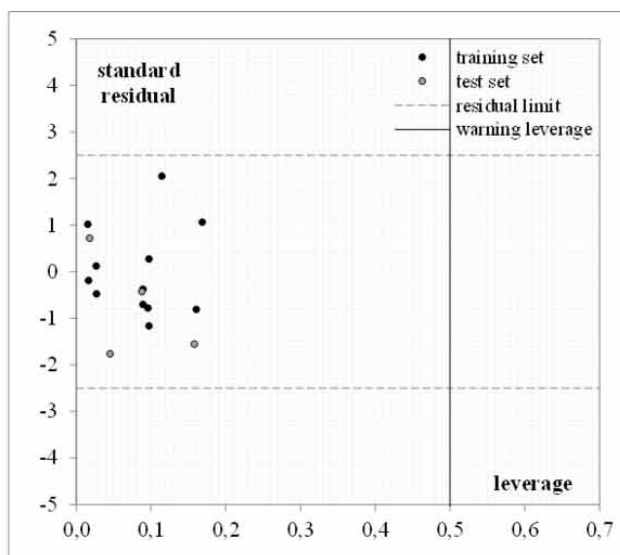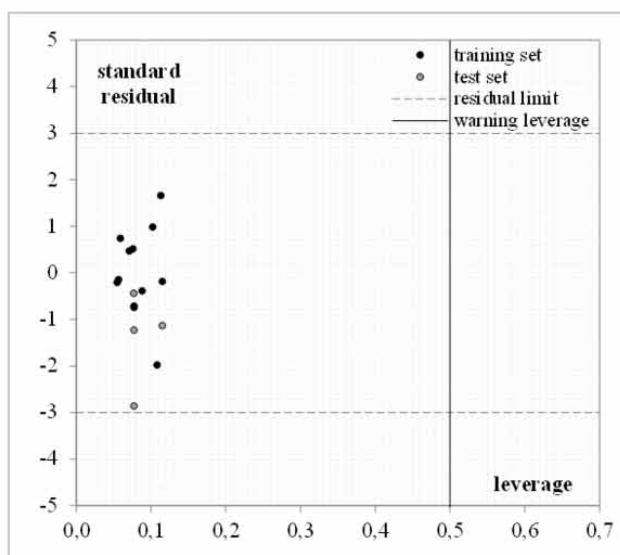
which is computed from the 14th order double stochastic graph-theoretic electronic-density matrix (duplex) of a molecular pseudo-graph, keeping all elements in the matrix form (entire matrix). It takes into account the Euclidean distance as invariant to characterize local properties of a molecule and the topological polar surface area as atom weight (Valdés-Martiní et al., 2017).

Although the chemical interpretation of any QuBiLs-MAS descriptor remains hidden behind the algebraic definition of this topological descriptor, it has a great importance for predicting pK_i in Equation 1. It is also possible to propose the following useful QSAR guide for the chemical synthesis of new cytochrome P450 3A4 inhibitors: according to Equation 1 and the sign of the regression

coefficient, the higher the value of d_1, the lower is the predicted pK_i (higher inhibitory potency) for a given anthocyanin structure falling within the AD of this model.

Considering now that a more rational partitioning method as BSM is employed, the following QSAR model is achieved:

$$pK_i = 0.45\,d_2 + 0.84 \tag{2}$$
$$N_{train} = 12\,,\ R^2_{train} = 0.99\,,\ RMS_{train} = 0.05$$
$$R^2_{loo} = 0.98\,,\ RMS_{loo} = 0.06\,,\ o3 = 0\,,\ R^2_{rand} = 0.91\,,\ RMS^{rand} = 0.13\,,\ R^2_m = 0.94$$
$$N_{test} = 4\,,\ R^2_{test} = 0.98\,,\ RMS_{test} = 0.09$$

It is to be appreciated that the statistical quality of Equation 2 is somewhat better than that provided by Equation 1. The pK_i predictions for Equation 2 are plotted in Figure 2B, while the applicability domain of this model is shown in Figure 3B, with all 4 test set molecules falling inside this AD.

In order to further validate the results obtained by Equations 1 and 2, additional validation parameters from the literature are provided in Table 1 for each QSAR equation, such as Q^2_{F1}, Q^2_{F2},

Table 1. Validation parameters for the two proposed QSAR models of mutagenicity

Equation	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC	R^2_m	R'^2_m	ΔR^2_m	$\overline{R^2_m}$
(1)	0.95	0.95	0.94	0.97	0.86	0.83	0.04	0.84
(2)	0.92	0.92	0.96	0.96	0.94	0.91	0.02	0.93

Q^2_{F3}, CCC, R^2_m, R'^2_m, ΔR^2_m and $\overline{R^2_m}$. It can be appreciated that satisfactory results are achieved.

The best selected descriptor d_2 belongs to the QuBiLs-MAS freeware, denoted by the symbol SD_B_AB_nCi_2_SS1_D_KA_psa-r_MAS. Such notation stands for the 1st order simple stochastic atom-based non-chiral and local (H bond donors, N and O atoms that have one bond with an H atom) bilinear index, which is computed from the 1st order simple stochastic graph-theoretic electronic-density matrix (duplex) of a molecular pseudo-graph, keeping all elements in the matrix form (entire matrix). It takes into account the standard deviation as invariant and both the topological polar surface area and refractivity as atom weights (Valdés-Martiní et al., 2017).

For the case of Equation 2, a QSAR guide is the following: the lower the value of d_1, the lower is the predicted pK_i (higher inhibitory potency) for a given anthocyanin structure falling within the defined AD.

The present QSAR results on the cytochrome P450 3A4 inhibition by anthocyanin derivatives, as expressed by Equations 1 and 2, compare fairly well with the 3D-QSAR models proposed in (Shityakov et al., 2014), where Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Index Analysis (CoMSIA) models are established. Table 2 includes the statistical quality of such models, where a better predictive capability is achieved in the test set by the QSAR models proposed in this study, either by using a random molecular partition of the dataset or a more rational partitioning in the BSM technique.

As it is well-known, the main advantage of the conformation-independent QSAR approach employed here, through the use of a great number of 0D-2D descriptors, is that it does not consider

Table 2. Statistical quality of 3D-QSAR models and the present study for predicting the inhibition constant

Model	R^2_{train}	RMS_{train}	R^2_{test}	RMS_{test}
CoMFA+	0.96	0.08	0.82	0.20
CoMSIA+	0.95	0.09	0.81	0.20
Equation 1	0.97	0.07	0.98	0.10
Equation 2	0.99	0.05	0.98	0.09

+Shityakov et al., 2014

the conformational representation of the chemical structures, relying only on their constitutional and topological representations (Duchowicz et al., 2015; Aranda et al., 2016; Ortiz et al., 2017; Fioressi et al., 2019).

In the analyzed antimutagenic dataset, there are only two anthocyanin derivatives with isomerism: Cyanidin 3-*O*-galactoside (3) and Cyanidin 3-*O*-glucoside (4); and Malvidin 3-*O*-galactoside (8) and Malvidin 3-*O*-glucoside (9). As can be noted from Table 3 in the Appendix, the four compounds are satisfactorily predicted and fall within the AD, and thereby one can conclude that no three-dimensional molecular descriptors are needed for the structural description in the present QSAR analysis.

Although Equations 1 and 2 have both an acceptable predictive capability on the training and test sets, the results found by these QSAR models have to be treated with certain caution. Only a few experimental data are used in the training and test sets, and a large number of molecular descriptors are analyzed for deriving the one-variable linear regressions, so it is true that the possibility of chance correlation cannot be completely discarded. In this work, the different evaluations of the model's performance have been successfully accomplished. As soon as available, more experimental data have to be included in future QSAR models for the inhibition of the mutagenic activity by anthocyanin derivatives.

4. CONCLUSION

The use of different freely available molecular descriptor calculation tools that prove to work well in a given dataset is of great value for the QSAR researcher. In addition, a proper partitioning of the molecular set, by considering the structure-activity relationships in the training and test sets, improves the statistical quality of the derived QSAR models.

In the present work, the analysis of 102,260 non-conformational molecular descriptors through univariable linear regression models achieve better predictive results on the test set than when more sophisticated techiques are used, like CoMSIA and CoMFA. Although a given structural molecular descriptor may not include chemical interpretation, because of its complex mathematical nature, the powerful of its prediction is that it enables to provide a QSAR guide for the search of new anthocyanin derivatives possessing high or low predicted mutagenic behavior.

FUNDING

AUTHORS' CONTRIBUTIONS

PRD and NAS established the QSAR models and drafted the manuscript. ABP participated in the design of the study and revised the manuscript. All authors read and approved the final manuscript for publication.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGMENT

DATA AVAILABLE ON REQUEST FROM THE AUTHORS

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

ACDLabs/ChemSketch. (n.d.). www.acdlabs.com

Bentz, E. N., Pomilio, A. B., & Lobayan, R. M. (2017). Donor-acceptor interactions as descriptors of the free radical scavenging ability of flavans and catechin. *Computational & Theoretical Chemistry*, *1110*, 14–24. doi:10.1016/j.comptc.2017.03.028

Chen, Z., Wang, C., Pan, Y., Gao, X., & Chen, H. (2018). Hypoglycemic and hypolipidemic effects of anthocyanins extract from black soybean seed coat in high fat diet and streptozotocin-induced diabetic mice. *Food & Function*, *9*(1), 426–439. doi:10.1039/C7FO00983F PMID:29220052

Cheynier, V., Tomas-Barberán, F. A., & Yoshida, K. (2015). Polyphenols: From plants to a variety of food and nonfood uses. *Journal of Agricultural and Food Chemistry*, *63*(35), 7589–7594. doi:10.1021/acs.jafc.5b01173 PMID:26281949

Chirico, N., & Gramatica, P. (2012). Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *Journal of Chemical Information and Modeling*, *52*(8), 2044–2058. doi:10.1021/ci300084j PMID:22721530

Dreiseitel, A., Schreier, P., Oehme, A., Locher, S., Hajak, G., & Sand, P. G. (2008). Anthocyanins and their metabolites are weak inhibitors of cytochrome P450 3A4. *Molecular Nutrition & Food Research*, *52*(12), 1428–1433. doi:10.1002/mnfr.200800043 PMID:18727015

Duchowicz, P. R., Szewczuk, N. A., & Pomilio, A. B. (2019). QSAR Studies of the Antioxidant Activity of Anthocyanins. *Journal of Food Science and Technology*, *56*(12), 5518–5530. doi:10.1007/s13197-019-04024-w PMID:31749500

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

Evteev, V. A., Shcherbak, N. P., & Kobliakov, V. A. (2006). A comparative study of induction regulation in cytochromes family 1 P450 in cell cultures at different stages of tumor transformation. *Tsitologiia*, *48*, 717–722. PMID:17089626

Fioressi, S. E., Bacelo, D. E., Rojas, C., Aranda, J. F., & Duchowicz, P. R. (2019). Conformation-independent Quantitative Structure-Property Relationships study on water solubility of pesticides. *Ecotoxicology and Environmental Safety*, *171*, 47–53. doi:10.1016/j.ecoenv.2018.12.056 PMID:30594756

Gasiorowski, K., Szyba, K., Brokos, B., Kołaczyńska, B., Jankowiak-Włodarczyk, M., & Oszmiański, J. (1997). Antimutagenic activity of anthocyanins isolated from Aronia melanocarpa fruits. *Cancer Letters*, *119*, 37–46. doi:10.1016/S0304-3835(97)00248-6 PMID:18372520

Gillam, E. M. J., & Hunter, D. J. B. (2007) Chemical defense and exploitation. Biotransformation of xenobiotics by Cytochrome P450 enzymes. In Metal Ions in Life Sciences: Vol. 3. The Ubiquitous Roles of Cytochrome P450 Proteins. West Sussex: Wiley and Sons.

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *Journal of Molecular Graphics & Modelling*, *20*(4), 269–276. doi:10.1016/S1093-3263(01)00123-1 PMID:11858635

Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., & Tong, W. et al. (2008). Mold^2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *Journal of Chemical Information and Modeling*, *48*(7), 1337–1344. doi:10.1021/ci800038f PMID:18564836

ISIDA/Fragmentor. (2017). *Laboratoire de Chémoinformatique, Chimie de la Matière Complexe* (SMS UMR 7140). Université de Strasbourg. http://complex-matter.unistra.fr/equipes-de-recherche/laboratoire-de-chemoinformatique/home

Kirton, S. B., Baxter, C. A., & Sutcliffe, M. J. (2002). Comparative modelling of cytochromes P450. *Advanced Drug Delivery Reviews*, *54*(3), 385–406. doi:10.1016/S0169-409X(02)00010-8 PMID:11922954

Langowski, J., & Long, A. (2002). Computer systems for the prediction of xenobiotic metabolism. *Advanced Drug Delivery Reviews*, *54*(3), 407–415. doi:10.1016/S0169-409X(02)00011-X PMID:11922955

Li, A., Yeo, K., Welty, D., & Rong, H. (2018). Development of guanfacine extended- release dosing strategies in children and adolescents with ADHD using a physiologically based pharmacokinetic model to predict drug–drug interactions with moderate CYP_3A_4 inhibitors or inducers. *Paediatric Drugs*, *20*(2), 181–194. doi:10.1007/s40272-017-0270-0 PMID:29098603

Lobayan, R. M., Jubert, A. H., Vitale, M. G., & Pomilio, A. B. (2009). Conformational and electronic (AIM/NBO) study of unsubstituted A-type dimeric proanthocyanidin. *Journal of Molecular Modeling*, *15*(5), 537–550. doi:10.1007/s00894-008-0389-6 PMID:19096885

Maestro 9.1. (n.d.). Schrödinger LLC.

Masand, V. H., & Rastija, V. (2017). PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. *Chemometrics and Intelligent Laboratory Systems*, *169*, 12–18. doi:10.1016/j.chemolab.2017.08.003

MATLAB R2013a. (n.d.). TheMathWorks, Inc. http://www.mathworks.com

Mercader, A. G., & Pomilio, A. B. (2011). *Biflavonoids: Occurrence, Structural Features and Bioactivity*. New York: Nova Science Publishers, Inc.

Miron, A., Aprotosoaie, A. C., Trifan, A., & Xiao, J. (2017). Flavonoids as modulators of metabolic enzymes and drug transporters. *Annals of the New York Academy of Sciences*, *1398*(1), 152–167. doi:10.1111/nyas.13384 PMID:28632894

Moon, Y. J., Wang, X., & Morris, M. E. (2006). Dietary flavonoids: Effects on xenobiotic and carcinogen metabolism. *Toxicology In Vitro*, *20*(2), 187–210. doi:10.1016/j.tiv.2005.06.048 PMID:16289744

Open Babel for Windows. (n.d.). http://openbabel.org/wiki/Category:Installation

Ortiz, E. V., Bennardi, D. O., Bacelo, D. E., Fioressi, S. E., & Duchowicz, P. R. (2017). The conformation-independent QSPR approach for predicting the oxidation rate constant of water micropollutants. *Environmental Science and Pollution Research International*, *24*(35), 27366–27375. doi:10.1007/s11356-017-0315-5 PMID:28975527

Ortiz de Montellano, P. R., & De Voss, J. J. (2005) Substrate oxidation by Cytochrome P450 enzymes. In Cytochrome P450: Structure, Mechanism, and Biochemistry (3rd ed.). Kluwer Academic/Plenum Publishers. doi:10.1007/0-387-27447-2_6

PaDEL 2.20 (Pharmaceutical Data Exploration Laboratory). (n.d.). http://www.yapcwsoft.com

Pomilio, A. B., & Mercader, A. G. (2018). Natural acylated anthocyanins and other related flavonoids: structure elucidation of Ipomoea cairica compounds and QSAR studies including multidrug resistance. In Studies in Natural Products Chemistry (Bioactive Natural Products). Elsevier Science Publishers. doi:10.1016/B978-0-444-64068-0.00009-7

Pomilio, A. B., Mercader, A. G., & Duchowicz, P. R. (2019) Quantitative structure–property relationship (QSPR) studies of alcoholic and non-alcoholic beverages, including wines, beers, and citrus juices. In Engineering Tools in the Beverage Industry: Vol. 3. The Science of Beverages. Woodhead Publishing, Elsevier Inc.

Putta, S., Yarla, N. S., Kumar, E. K., Lakkappa, D. B., Kamal, M. A., Scotti, L., & Aliev, G. et al. (2018). Preventive and therapeutic potentials of anthocyanins in diabetes and associated complications. *Current Medicinal Chemistry*, *25*(39), 5347–5371. doi:10.2174/0929867325666171206101945 PMID:29210634

Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015). 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 PMID:26521096

Roy, K., Chakraborty, P., Mitra, I., Ojha, P. K., Kar, S., & Das, R. N. (2013). Some case studies on application of "r2m" metrics for judging quality of Quantitative Structure–Activity Relationship predictions: Emphasis on scaling of response data. *Journal of Computational Chemistry*, *34*(12), 1071–1082. doi:10.1002/jcc.23231 PMID:23299630

Roy, P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules (Basel, Switzerland)*, *14*(5), 1660–1701. doi:10.3390/molecules14051660 PMID:19471190

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

Sánchez-Jiménez, F., Ruiz-Pérez, M. V., Urdiales, J. L., & Medina, M. A. (2013). Pharmacological potential of biogenic amine-polyamine interplay beyond neurotransmission. *British Journal of Pharmacology*, *170*(1), 4–16. doi:10.1111/bph.12109 PMID:23347064

Selvaraj, C., Tripathi, S. K., Reddy, K. K., & Singh, S. K. (2011). Tool development for prediction of pIC_{50} values from the IC_{50} values-a pIC_{50} value calculator. *Current Trends in Biotechnology and Pharmacy*, *5*, 1104–1109.

Shityakov, S., Puskás, I., Roewer, N., Förster, C., & Broscheit, J. (2014). Three-dimensional quantitative structure-activity relationship and docking studies in a series of anthocyanin derivatives as cytochrome P450 3A4 inhibitors. *Advances and Applications in Bioinformatics and Chemistry: AABC*, *7*, 11–21. doi:10.2147/AABC.S56478 PMID:24741320

Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics* (Vol. 41). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9783527628766

Tute, M. S. (1971). *History and Objectives of Quantitative Drug Design in Advances in Drug Research*. London: Academic Press.

Valdés-Martiní, J. R., Marrero-Ponce, Y., García-Jacas, C. R., Martinez-Mayorga, K., Barigye, S. J., Vaz d'Almeida, Y. S., & Morell, C. A. et al. (2017). QuBiLS-MAS, open sourcemulti-platform software foratom- and bond-based topological (2D) and chiral (2.5D) algebraic molecular descriptors computations. *Journal of Cheminformatics*, *9*(1), 35–61. doi:10.1186/s13321-017-0211-5 PMID:29086120

Yamazaki, H., Shaw, P. M., Guengerich, F. P., & Shimada, T. (1998). Roles of cytochromes P450 1A2 and 3A4 in the oxidation of estradiol and estrone in human liver microsomes. *Chemical Research in Toxicology*, *11*(6), 659–665. doi:10.1021/tx970217f PMID:9625734

Zanger, U. M., & Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*, *138*(1), 103–141. doi:10.1016/j.pharmthera.2012.12.007 PMID:23333322

APPENDIX: SUPPLEMENTAL DATA

Table 3. Experimental and predicted inhibitory activities for 16 anthocyanin derivatives. Test set compounds are denoted with ^.

ID	Chemical Name	pK_i exp	Equation 1	Equation 2	d_1	d_2
1	Cyanidin	4.680	4.670	4.627	225.4746	8.4052
2	Cyanidin 3,5-O-diglucoside	3.760	3.675	3.720	564.0119	6.3896
3	Cyanidin 3-O-galactoside	4.095	4.073	4.134	428.5763	7.3079
4	Cyanidin 3-O-glucoside	3.980	4.073	4.134^	428.5763	7.3079
5	Cyanidin 3-O-rutinoside	4.020	4.083	3.995	425.4125	6.9995
6	Delphinidin 3-O-glucoside	4.130	3.966	4.102	465.0745	7.2385
7	Malvidin 3,5-O-diglucoside	3.650	3.715	3.658	550.5767	6.2505
8	Malvidin 3-O-galactoside	4.070	4.126	4.136^	410.5934	7.3134
9	Malvidin 3-O-glucoside	4.096	4.126	4.136	410.5934	7.3134
10	Pelargonidin	4.910	4.829	4.821	171.6725	8.8351
11	Peonidin	4.795	4.810	4.856^	177.9036	8.9132
12	Petunidin	4.630	4.668	4.736	226.1704	8.6476
13	Delphinidin	4.333	4.474^	4.354	292.2486	7.7974
14	Malvidin	4.850	4.793^	4.860	183.8355	8.9224
15	Pelargonidin 3,5-O-diglucoside	3.603	3.728^	3.614	546.2345	6.1537
16	Peonidin-3-O-glucoside	4.101	4.135^	4.124^	407.5496	7.2878

Nicolas Alejandro Szewczuk is a doctoral fellow of the National Council of Scientific and Technical Research (CONICET). Degree in Chemistry from the National University of La Plata (Argentina), dedicated to the theoretical-structural study of compounds belonging to the family of flavonoids.

Pablo R. Duchowicz (PhD) studied Physical-Chemistry from 1996-2003 at the Faculty of Exact Sciences, Chemistry Department of La Plata National University (UNLP), Buenos Aires, Argentina. His Ph.D. work on 'Physicochemical and Biological Applications of the QSAR/QSPR Theory', was performed at the Research Institute of Theoretical and Applied Physical-Chemistry (INIFTA) located at La Plata, under the supervision of Profs. Dr. Eduardo A. Castro and Dr. Francisco M. Fernández, from 2003-2005. In 2006 he obtained a Post-doctoral scholarship to work on 'ab initio Direct Kinetics and Molecular Dynamics Studies for Halogenated Germanes and Related Species' at INIFTA, under the supervision of Prof Dr. Carlos J. Cobos. Since 2007, he is a member of the Scientific Researcher Career of the National Research Council of Argentina, performing his research work mainly on QSAR/QSPR studies at INIFTA.

Prof. Alicia B. Pomilio (PhD) has a post Doctorate Degree (post-Ph.D.): Organisch-Chemisches Institut der Universität Heidelberg, Germany (as Research Fellowship of the Alexander von Humboldt-Stiftung, German International Research Foundation). Researcher in Germany. Her Ph.D. work on Organic Chemistry, "Structures of novel simple and acylated anthocyanins", was developed at the Facultad de Ciencias Exactas y Naturales (FCEN), Universidad de Buenos Aires (UBA). At present, Dra. Pomilio is Professor of the University of Buenos Aires. Founder and Director of the Research Institute PROPLAME of CONICET (currently INMIBO) (FCEN, UBA) as Senior Research Member of CONICET. Dra Pomilio has performed as PhD dissertation director of 20 Ph. D. Theses, 3 Master research works, and 18 Bachelor Theses (Universidad de Buenos Aires, Universidad Nacional de La Plata, Universidad Nacional del Sur y Universidad Nacional de Mar del Plata). Total: 41 (approved with maximum score). About 200 research articles in international scientific journals; 9 Books; 7 Book chapters. Some 260 Symposium presentations; Plenary Conferences and Workshops at Research Centers of American, European and US Countries. Also conferences in national and international organizations. A number of 18 Awards for scientific accomplishments.