



# QSAR studies of the antioxidant activity of anthocyanins

Pablo R. Duchowicz<sup>1</sup> · Nicolás A. Szewczuk<sup>1</sup> · Alicia B. Pomilio<sup>2</sup>

Revised: 8 August 2019 / Accepted: 12 August 2019 / Published online: 17 August 2019  
© Association of Food Scientists & Technologists (India) 2019

**Abstract** Through experimental information available from antioxidant assays of seventeen anthocyanins, and six common anthocyanidins, quantitative structure–activity relationships (QSAR) have been established in the present work. The antioxidant bioactivity has been predicted in three different lipid environments: emulsified and bulk oil (methyl linoleate) (in vitro tests) at concentrations of 50 and 250  $\mu\text{M}$ , and 50  $\mu\text{M}$  of the inhibitor, respectively, and in human LDL (low-density lipoprotein; “bad cholesterol”) (ex vivo test) at concentrations of 2.5, 10, and 25  $\mu\text{M}$  of the inhibitor. Radical scavenging activity was predicted in the assay with the 1,1-diphenyl-2-picrylhydrazyl radical (DPPH $\cdot$ ). The QSAR models developed for each test and concentration used allowed to obtain prospective information on the constitutional and topological molecular characteristics for anthocyanin/anthocyanidin compounds. Therefore, the antioxidant activity was predicted for twenty-one compounds with unknown experimental values, leading for some of them to a favorable predicted bioactivity.

**Keywords** Anthocyanins · Antioxidant activity · Quantitative structure–activity relationships · Molecular descriptors

## Introduction

Anthocyanins are natural hydrosoluble dyes occurring as *O*-glycosides in flowers, fruits, leaves, bracts, and stems that belong to the family of flavonoids (Harborne and Williams 2000; He and Giusti 2010; Du et al. 2018; Fujiwara et al. 2018). These compounds are constituted by an aglycone (anthocyanidin) that contains a positive charge on the oxygen atom (oxonium salt), commonly known as a flavylium salt or a 2-phenyl-chromilium ion. Several 3,5,7-trisubstituted hydroxylated anthocyanidins have been isolated and further identified, 5- and 7-hydroxyls being sometimes methylated. Also 6- or 8-hydroxylated anthocyanidins, and 3- and 5-deoxyanthocyanidins have been reported (Sousa et al. 2016).

The structural variability of anthocyanins is due to the hydroxylation and methoxylation patterns of the aglycones, and also to the glycosylation pattern, including sugar acylation (Sakuta 2014). Acylated anthocyanins are of interest due to their restricted taxonomic distribution, bioactivity, and biogenetic relationship with sugar esters of cinnamic acid derivatives (Pomilio and Mercader 2017; Zhang et al. 2019).

Anthocyanins (*O*-glycosides) and their aglycones (anthocyanidins) show several bioactivities of therapeutic value against chronic diseases, such as anti-inflammatory, antiangiogenic, antimutagenic, antiproliferative, antioxidant, anticarcinogenic, cardioprotective, estrogenic, chondroprotective, and anti-neurodegenerative properties (Bonesi et al. 2019; Li et al. 2019; Tsakiroglou et al. 2019;

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13197-019-04024-w>) contains supplementary material, which is available to authorized users.

✉ Pablo R. Duchowicz  
pablotucho@gmail.com

<sup>1</sup> Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas (INIFTA), CONICET, Universidad Nacional de La Plata (UNLP), Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina

<sup>2</sup> Laboratorio de Química y Bioquímica Estructural, Departamento de Bioquímica Clínica, Hospital de Clínicas “José de San Martín”, Universidad de Buenos Aires, Av. Córdoba 2351, C1120AAF Buenos Aires, Argentina

Ullah et al. 2019; Vishnu et al. 2019; Wongwichai et al. 2019; Zhang et al. 2019). Mechanisms of action have been reported, including effects on gene and protein expression as well as interactions with micro-RNAs (miRNAs) and cell-signaling pathways (Khan et al. 2019; Krga and Milenkovic 2019; Ma and Ning 2019; de Sousa Moraes et al. 2019). Target cells of anthocyanins seem to be adipocytes, endothelial cells, inflammatory cells, hepatocytes, and intestinal cells (Jiang et al. 2019).

In particular, antioxidant activities (Kaurinovic and Vastag 2019) explain the value of some functional foods in human health, and the importance of nutraceuticals to prevent diseases, such as cardiovascular, and inflammatory diseases, including diabetes and metabolic syndrome. Plant-derived antioxidants have been shown to function as singlet and triplet oxygen quenchers, free radical scavengers, peroxide decomposers, metal chelators, enzyme inhibitors and synergists (Jaganath and Crozier 2010).

It is worth noting that free radicals are highly reactive and, therefore, very unstable, with a very short life time [minutes for hydrogen peroxide ( $H_2O_2$ ), and organic hydroperoxides; seconds for the peroxy radical; milliseconds for peroxy nitrite; microseconds for superoxide anion, singlet oxygen, alkoxy radicals; and nanoseconds for the hydroxyl radical]. The presence of free radicals in living organisms can only be explained by compartmentalization at cellular level (Thankam Finosh and Jayabalan 2013). Therefore, the main product of cell metabolism is  $H_2O_2$ , whose formation is thermodynamically favorable because of a positive redox potential. Anthocyanins are known to remove  $H_2O_2$  from cells that mainly occurs by the Fenton reaction. It takes place through the hydroxyalkene group of the anthocyanin structures via the mechanism proposed for this reaction (Vitale et al. 2016). Hence, a large part of their bioactivities can be explained by this important function.

The potential chemopreventive use of anthocyanins from functional foods or as supplements, together with the possibility of using anthocyanins as natural food colorants led us to perform studies of quantitative structure–activity relationships (QSAR). In this way, mathematical models with predictive power were obtained by associating the molecular structure, encoded within the so-called molecular descriptors, with the biological activity of interest, e.g., the antioxidant activity.

Therefore, the aim of the present work was to obtain simple, novel, and easily interpretable QSAR models to predict the antioxidant properties of twenty-three anthocyanins (aglycones, *O*-glycosides, and *O*-acylglycosides) by evaluating the free radical scavenging activity, and percentage inhibition of lipid oxidation.

Structural features such as the extent of hydroxylation and methoxylation of ring B, and the type of glycosylation/

acylglycosylation of the anthocyanins have been previously reported to modify the stability, bioavailability, and antioxidant/anti-inflammatory effects of anthocyanins (Kähkönen and Heinonen 2003; Miguel 2011). These retrospective analyses based on experimental values are very useful, but we encouraged the present QSAR study due to the need to deepen into the whole molecular structure. QSAR developments involved prospective information based on the constitutional and topological molecular characteristics of the anthocyanins. Therefore, the antioxidant activity was analyzed and quantified for the predictive design of new antioxidants.

Finally, a practical application of the established QSAR models was performed. Acceptable predicted antioxidant bioactivities were obtained for twenty-one anthocyanins/anthocyanidins with unknown experimental values, most of them being considered as reliable predictions according to their leverage values.

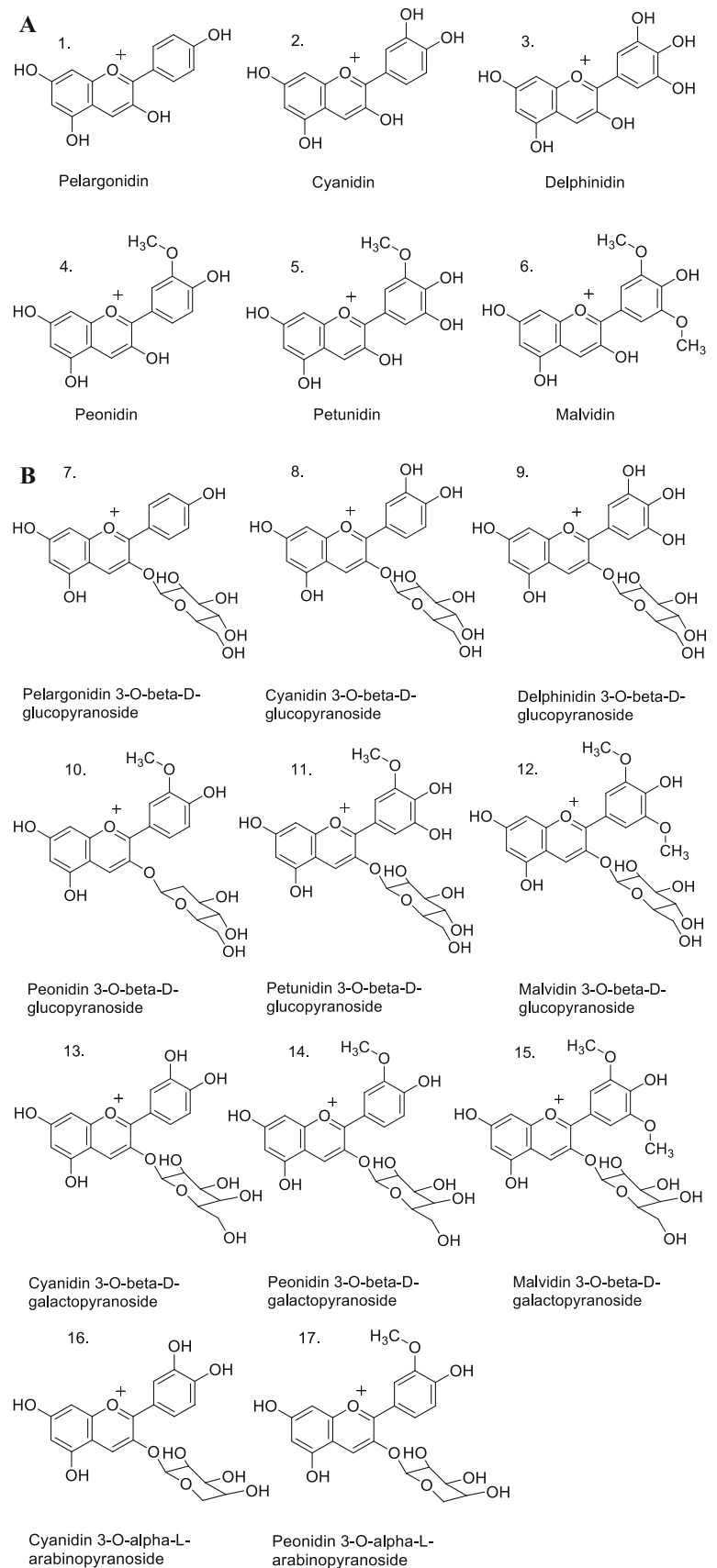
## Materials and methods

### Anthocyanins selected for this study

Seventeen anthocyanins and six anthocyanidins were selected, from which experimental information on antioxidant tests was available (Kähkönen and Heinonen 2003). The chemical structures of the selected anthocyanins, with increasing complexity, corresponded to:

- (a) *Anthocyanidins* (without glycosidation), such as: pelargonidin (1), cyanidin (2), delphinidin (3), peonidin (4), petunidin (5), and malvidin (6) (Fig. 1a).
- (b) *Anthocyanins: 3-O-monoglycosides of anthocyanidins*, with monosaccharides (glucose, galactose, and arabinose) in the 3-position of the aglycones, such as: pelargonidin 3-*O*- $\beta$ -D-glucopyranoside (7), cyanidin 3-*O*- $\beta$ -D-glucopyranoside (8), delphinidin 3-*O*- $\beta$ -D-glucopyranoside (9), peonidin 3-*O*- $\beta$ -D-glucopyranoside (10), petunidin 3-*O*- $\beta$ -D-glucopyranoside (11), malvidin 3-*O*- $\beta$ -D-glucopyranoside (12); cyanidin 3-*O*- $\beta$ -D-galactopyranoside (13), peonidin 3-*O*- $\beta$ -D-galactopyranoside (14), malvidin 3-*O*- $\beta$ -D-galactopyranoside (15); cyanidin 3-*O*- $\alpha$ -L-arabinopyranoside (16), peonidin 3-*O*- $\alpha$ -L-arabinopyranoside (17) (Fig. 1b).
- (c) *Anthocyanins: 3-O-diglycosides of anthocyanidins*, with disaccharide (rutin = disaccharide 6-*O*- $\alpha$ -L-rhamnopyranosyl- $\beta$ -D-glucopyranose), in the 3-position of the aglycone, such as: cyanidin 3-*O*- $\beta$ -D-rutinoside (18), delphinidin 3-*O*- $\beta$ -D-rutinoside (19);

**Fig. 1** Structures of: **A** anthocyanidins **1–6**; **B** monoglycosylated anthocyanins **7–17**; **C** diglycosylated anthocyanins **18–21**; **D** triglycosylated anthocyanin **22**, and acylated triglycosylated anthocyanin **23**



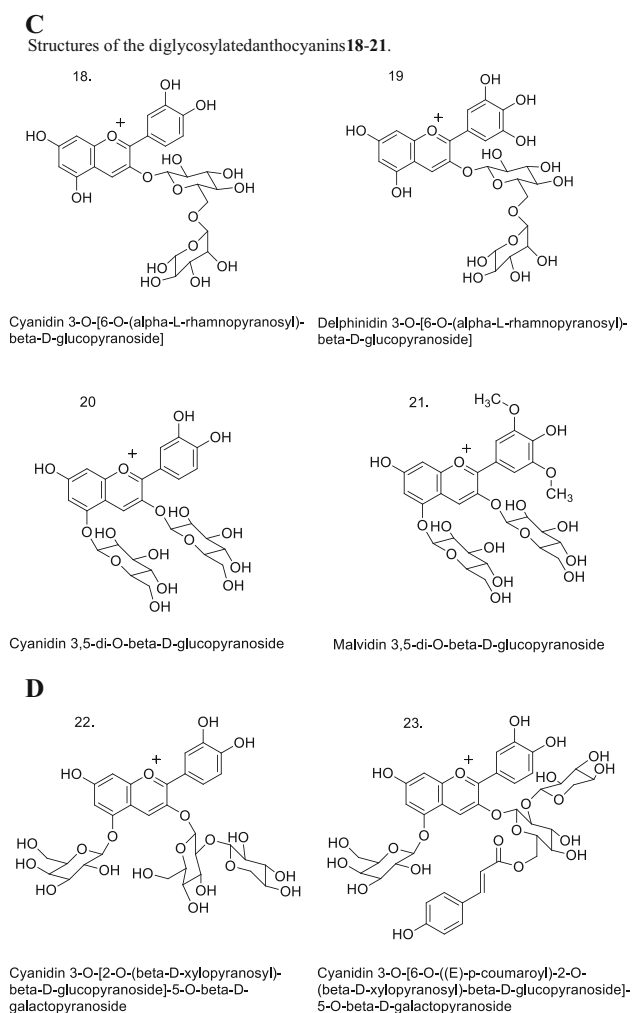


Fig. 1 continued

- (d) *Anthocyanins: 3,5-O-diglycosides of anthocyanidins*, with monosaccharides (glucose) in the 3- and 5-positions of the aglycone, such as: cyanidin 3,5-di-O-β-D-glucopyranoside (**20**), malvidin 3,5-di-O-β-D-glucopyranoside (**21**) (Fig. 1c).
- (e) *Anthocyanin: 3-O-diglycoside-5-O-monoglycoside of anthocyanidin*, with disaccharide (xylosylglucoside = sambubioside = disaccharide 2-O-β-D-xylopyranosyl-β-D-glucopyranose) in the 3-position, and monosaccharide (galactose) in the 5-position of the anthocyanidin, such as: cyanidin 3-O-β-D-sambubioside-5-O-β-D-galactopyranoside (**22**);
- (f) *Monoacylated anthocyanin: 3-O-monoacyldiglycoside-5-O-monoglycoside of anthocyanidin*, with monoacylated disaccharide (coumaroyl-xylosylglucoside = E-p-coumaroylsambubioside = 6''-O-E-p-coumaroylsambubioside) in the 3-position, and monosaccharide (galactose) in the 5-position of the anthocyanidin; the acyl corresponds to a derivative

of cinnamic acid as p-coumaroyl, such as: cyanidin 3-O-β-D-(6''-O-E-p-coumaroyl)sambubioside-5-O-β-D-galactopyranoside (**23**) (Fig. 1d).

Well-known antioxidants were also tested for comparative purposes (Kähkönen and Heinonen 2003): ascorbic acid, α-tocopherol, and Trolox, three phenolic acids: gallic, caffeic, and chlorogenic acids, and three flavonoids: quercetin, rutin, and (+)-catechin (Bentz et al. 2017).

### Antioxidant activity evaluation

The antioxidant activity of anthocyanidins and their glycosides was evaluated in three different lipid environments: emulsified and bulk oil (methyl linoleate: MeLo) (in vitro tests) at concentrations of 50 and 250 μM, and 50 μM of the inhibitor, respectively, and in human low-density lipoprotein (LDL; “bad cholesterol”) (ex vivo test) at concentrations of 2.5, 10, and 25 μM of the inhibitor. In addition, the radical scavenging activity of the compounds was studied against the 1,1-diphenyl-2-picrylhydrazyl radical (DPPH·). The evaluation methodologies have been previously reported (Kähkönen and Heinonen 2003).

### QSPR studies on the anthocyanins

The structures of the anthocyanins were represented with conformation-independent molecular descriptors in order to consider their constitutional and topological nature. The molecules were drawn in the MDL mol (V2000) format in ACD/ChemSketch (2016), PaDEL (2016), Mold<sup>2</sup> (Hong et al. 2008), EPI Suite (2018), RECON (Lavine et al. 2003), and QuBiLS-MAS (Valdes-Martini et al. 2012) softwares were used to calculate 46,994 linearly independent descriptors. All these programs are free and open access.

### Data set of anthocyanins

The molecular set was divided into a training set (train) to select the molecular descriptors and adjustable parameters of the multivariable linear regression (MLR), and an external testset (test) to determine the predictive capacity of the obtained model. The test set compounds were selected by approximately spanning the experimental bioactivity interval.

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

Table S1 includes the experimental antiradical and antioxidant activities in scaled and logarithmic units

**Table 1** A brief description for non-conformational descriptors appearing in this QSAR study, obtained from PaDEL (2016), Mold<sup>2</sup> (Hong et al. 2008), and QuBiLS-MAS (Valdes-Martini et al. 2012) programs

Type of descriptor	Symbol	Description
QuBiLS-MAS	<i>qub1, qub2, qub3, qub4, qub5, qub6, qub7</i> <sup>a</sup>	Mathematical indexes obtained from quadratic, bilinear, and linear-type algebraic forms, based on matrices of the Graph theory and atomic weights
Electrotopological	<i>minHBint4</i>	Minimum E-state descriptor of strength for potential hydrogen bonds of path length 4
Mold <sup>2</sup>	<i>D284</i>	structural information content index of order-3
	<i>D714</i>	count of CH <sub>3</sub> R and CH <sub>4</sub> groups
2D Autocorrelation	<i>D458</i>	Geary autocorrelation—lag 4, weighed by atomic van der Waals volumes
	<i>GATS6p</i>	Geary autocorrelation—lag 6, weighed by atomic polarizabilities
	<i>ATSC8e</i>	Centered Broto-Moreau autocorrelation—lag 8/weighted by atomic Sanderson electronegativities
Baryszmatrix	<i>VR1.Dzm</i>	Randic-type eigenvector-based index from Barysz matrix/weighted by mass
	<i>VR2.Dzs</i>	Normalized Randic-type eigenvector-based index from Barysz matrix/weighted by I-state

<sup>a</sup>*qub1* = SD\_B\_AB\_nCi\_2\_SS8\_T\_KA\_psa-p\_MAS; *qub2* = SD\_B\_AB\_nCi\_2\_SS8\_T\_KA\_psa-e\_MAS; *qub3* = AM\_B\_AB\_nCi\_2\_DS3\_M\_KA\_v-c\_MAS; *qub4* = SD\_B\_AB\_nCi\_2\_SS7\_D\_KA\_psa-v\_MAS; *qub5* = AM\_B\_AB\_nCi\_2\_SS15\_T\_KA\_psa-r\_MAS; *qub6* = SD\_B\_AB\_nCi\_2\_SS8\_A\_KA\_psa-p\_MAS; *qub7* = AM\_Q\_AB\_nCi\_2\_DS5\_T\_KA\_psa\_MAS

[log(I% + 100)] of the selected anthocyanidins and anthocyanins (Kähkönen and Heinonen 2003). The test set compounds are indicated with ^ for each activity analyzed.

The selected data reduction technique will make the QSAR models differ in their accuracy, complexity, and predictive capacity. The most direct strategy is MLR, where the variables of the linear equation are a few descriptors that best predict the activity. As it is known, the selection of descriptors is not easy, that is why the QSAR models were established with the “Replacement Method” (RM) (Duchowicz et al. 2006). This technique consists of a mathematical algorithm designed in our research group, which allows selecting the best descriptors from a large number of them. This selection is performed by minimizing the square root of the mean square error of the training set ( $RMS_{train}$ ).

The approximate solutions achieved with RM sometimes agree with the exact solutions (complete analysis of all possible variable combinations). In other cases they are very close (Duchowicz et al. 2006; Mercader et al. 2010). All Matlab algorithms used were developed in our research group, and are available by request.

Once the best solution was identified, the techniques of the external validation set, cross validation, and applicability domain (Roy et al. 2015) allowed determining whether the model was acceptable.

## Results and discussion

Quantitative structure–activity relationships (QSAR) models were established in order to predict the inhibition percentage of oxidation product formation (I%) when the inhibitors were added at different concentrations. The

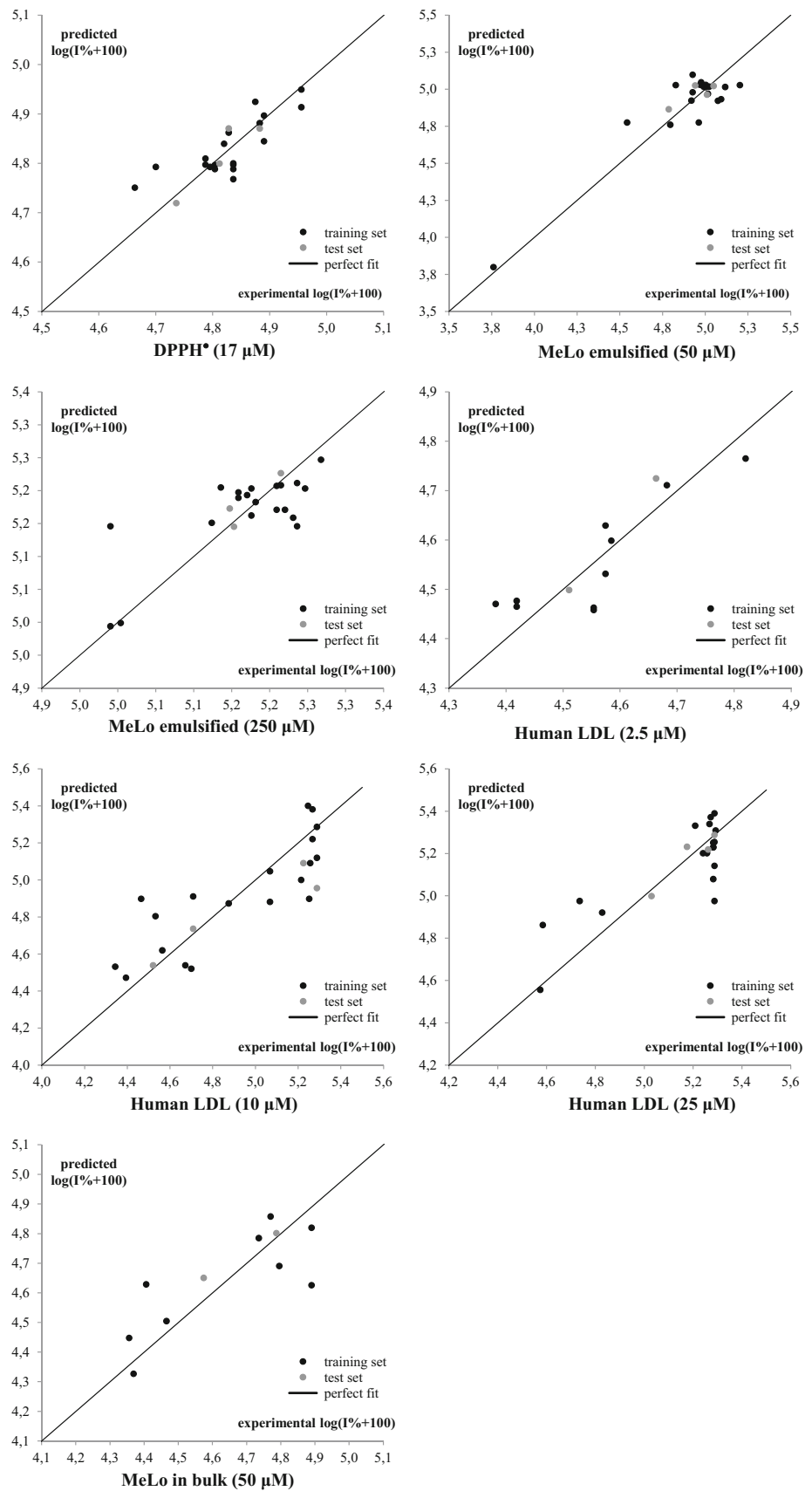
structures of the anthocyanins were represented with conformation-independent molecular descriptors in order to take into account their constitutional and topological nature.

The main advantage of not considering molecular conformations is that the only experimental information required to establish the QSAR models is the biological activity that is analyzed (Duchowicz 2018), in our case the antioxidant activity. No additional experimental information was required, such as that provided, for example, by X-ray structural crystallographic data for some anthocyanin conformation in its molecular environment, data that are often unknown. In this way, a QSAR model was established based on the minimum essential experimental information (experimental biological activity, I%). Therefore, conformation-independent QSAR studies were very useful for application to any chemical system of interest.

The most representative non-conformational molecular descriptors found in this work for the studied bioactivities are briefly described in Table 1, being involved in the different quantitative structure–antioxidant activity relationships of the 23 anthocyanins analyzed. These descriptors were selected from 46,994 linearly independent descriptors available at PaDEL (2016), Mold<sup>2</sup> (Hong et al., 2008), EPI Suite (2018), RECON (Lavine et al. 2003), and QuBiLS-MAS (Valdes-Martini et al. 2012) programs. The numerical values of such descriptors are displayed in Table S2.

The best QSAR linear regressions, which predict the inhibition percentage of oxidation products formation (I%) when the inhibitors are added at different concentrations are shown as Eqs. (1)–(7). Each model had an acceptable number of molecules to number of descriptors ratio ( $N_{train}/d \geq 6$ ) in order to avoid data overfitting.

**Fig. 2** Predictions of the antioxidant activity according to QSAR models of Eqs. (1)–(7)



DPPH<sup>•</sup> (17 μM):

$$\begin{aligned} \log(I\% + 100) &= -2.27GATS6p - 0.070qub1 + 8.81 \\ N_{train} &= 19, R_{train}^2 = 0.63, RMS_{train} = 0.04, o_2 = 0, \\ R_{ij}^{2max} &= 0.36N_{test} = 4, R_{test}^2 = 0.85, RMS_{test} = 0.02, \\ R_{loo}^2 &= 0.51, RMS_{loo} = 0.05 \end{aligned} \quad (1)$$

MeLo emulsion (50 μM):

$$\begin{aligned} \log(I\% + 100) &= -1.35D714 + 2.88 \cdot 10^{-4} VR1.Dzm \\ &\quad - 0.024VR2.Dzs + 5.17 \\ N_{train} &= 19, R_{train}^2 = 0.85, RMS_{train} = 0.12, o_2 = 0, \\ R_{ij}^{2max} &= 0.48N_{test} = 4, R_{test}^2 = 0.67, RMS_{test} = 0.06, \\ R_{loo}^2 &= 0.59, RMS_{loo} = 0.19 \end{aligned} \quad (2)$$

MeLo emulsion (250 μM):

$$\begin{aligned} \log(I\% + 100) &= 5.87 \cdot 10^{-5} VR1.Dzm - 0.057qub2 \\ &\quad - 9.47qub3 + 5.78 \\ N_{train} &= 20, R_{train}^2 = 0.59, RMS_{train} = 0.05, o_3 = 0, \\ R_{ij}^{2max} &= 0.46N_{test} = 3, R_{test}^2 = 0.83, RMS_{test} = 0.02, \\ R_{loo}^2 &= 0.52, RMS_{loo} = 0.06 \end{aligned} \quad (3)$$

Human LDL (2.5 μM):

$$\begin{aligned} \log(I\% + 100) &= -0.11qub4 + 9.27 \\ N_{train} &= 10, R_{train}^2 = 0.74, RMS_{train} = 0.06, o_2 = 0 \\ N_{test} &= 2, R_{test}^2 = 1.00, RMS_{test} = 0.04, R_{loo}^2 = 0.59, \\ RMS_{loo} &= 0.08 \end{aligned} \quad (4)$$

Human LDL (10 μM):

$$\begin{aligned} \log(I\% + 100) &= 0.66ATSC8e + 0.14qub5 - 0.64qub6 \\ &\quad + 13.80N_{train} = 19, R_{train}^2 = 0.69, RMS_{train} = 0.19, \\ o_{2.5} &= 0, R_{ij}^{2max} = 0.49 \\ N_{test} &= 4, R_{test}^2 = 0.90, RMS_{test} = 0.18, R_{loo}^2 = 0.58, \\ RMS_{loo} &= 0.22 \end{aligned} \quad (5)$$

Human LDL (25 μM):

$$\begin{aligned} \log(I\% + 100) &= 1.26D284 + 0.43minHBint4 \\ &\quad + 0.021qub7 - 3.00 \\ N_{train} &= 19, R_{train}^2 = 0.69, RMS_{train} = 0.14, o_{2.5} = 0, \\ R_{ij}^{2max} &= 0.84N_{test} = 4, R_{test}^2 = 0.88, RMS_{test} = 0.04, \\ R_{loo}^2 &= 0.54, RMS_{loo} = 0.17 \end{aligned} \quad (6)$$

MeLo in bulk (50 μM):

$$\begin{aligned} \log(I\% + 100) &= 2.70D458 + 3.40 \\ N_{train} &= 9, R_{train}^2 = 0.62, RMS_{train} = 0.13, o_2 = 0 \\ N_{test} &= 2, R_{test}^2 = 1.00, RMS_{test} = 0.054, R_{loo}^2 = 0.50, \\ RMS_{loo} &= 0.16 \end{aligned} \quad (7)$$

In these equations,  $R$  is the linear correlation coefficient;  $o \cdot m$  is the number of calibration molecules that have a residual greater than the  $m \cdot S_{train}$  value, where  $S_{train}$  is the standard deviation for the training set; and  $R_{ij}^{max}$  is the maximum intercorrelation between descriptor pairs.

The molecular descriptors involved in Eqs. (1)–(7) tended to accurately predict the antioxidant potential, as shown in Fig. 2. A linear trend was observed when plotting the predicted antioxidant activity as a function of the experimental activity. The maximum intercorrelation between pairs of descriptors ( $R_{ij}^{max}$ ) was not a serious problem in Eqs. (1)–(7), suggesting that each descriptor involved in the linear regression included non-overlapping structural information.

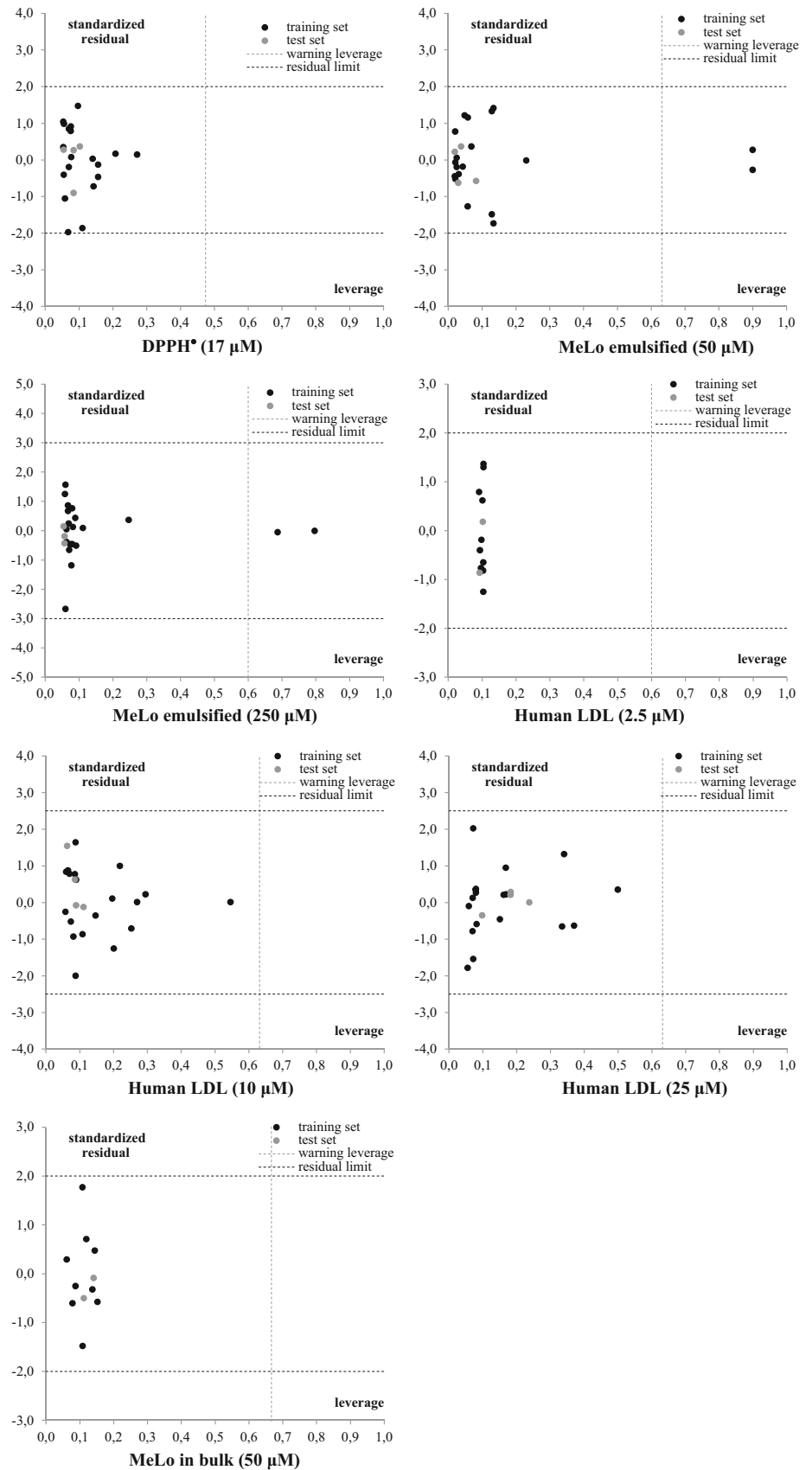
The residual scatter plots of the QSAR models showed that the data tended to be randomly distributed around the zero line, thus suggesting the absence of systematic errors during modeling (Fig. S1). It is worth to mention that it was not possible to obtain an absolute conclusion of the pattern shown by such graphs, due to the small number of observations involved ( $N = 23$ ). In addition, although some anthocyanins of these graphs exhibited high residual (difference between the value of the experimental activity, and the predicted one), all structures were included, deciding not to remove any of them of the model in order to be able to predict and analyze them in the present dataset. Finally, none of the established models showed molecules with residues greater than the  $3 \cdot S_{train}$  value, revealing the absence of molecules with anomalous behavior in terms of their structure–activity relationship.

The predictive capacity of these models for antioxidant activity was satisfactory, as suggested by the  $RMS_{test}$  parameter, which was lower or comparable to  $RMS_{train}$  in Eqs. (1)–(7). Therefore, the anthocyanins of the external

**Table 2** Different validation parameters for Eqs. (1)–(7)

Equations	$Q_{F1}^2$	$Q_{F2}^2$	$Q_{F3}^2$	CCC	$R_m^2$
(1)	0.80	0.78	0.88	0.91	0.63
(2)	0.67	0.60	0.96	0.73	0.55
(3)	0.71	0.70	0.96	0.87	0.68
(4)	0.72	0.67	0.88	0.90	0.53
(5)	0.70	0.70	0.72	0.80	0.64
(6)	0.87	0.85	0.98	0.93	0.74
(7)	0.79	0.74	0.94	0.85	0.70

**Fig. 3** Standardized residuals as function of the influence parameter





test set were adequately predicted with the QSAR models that were obtained only on the basis of the calibration set. Here, it should be mentioned that  $R_{test}$  was not a significant validation parameter in the present analysis due to the small number of validation molecules used ( $N_{test} \leq 5$ ).

Apart from using a test set, each model was theoretically validated by means of ‘Leave-One-Out’ (loo) cross-validation. The  $R_{loo}$  and  $RMS_{loo}$  parameters revealed that the Eqs. (1)–(7) were stable when any molecule was removed from the training set, and was predicted with the recalculated model with the remaining calibration molecules (using the same model descriptors).

Further validation parameters from the literature were provided in Table 2 for each QSAR equation, achieving moderate results. However, it was noted that these parameters (e.g.,  $Q_{F1}^2$ ,  $Q_{F2}^2$ ,  $Q_{F3}^2$ ,  $CCC$ , and  $R_m^2$ ) would have worked better if a greater number of data points were included in the QSAR study. A better behavior would also have been obtained if these data points uniformly covered the range of experimental activity, which was not the case in the bioactivities of the present work.

As it is known, a predictive QSAR model is only able to predict molecules that belong to its applicability domain (AD), so that the activity predicted by the model is not the result of extrapolation (unreliable prediction) (Roy et al. 2015). In the present work, the AD was evaluated by the influence parameter ( $h$ ): the prediction of a test set molecule was reliable provided when its influence was less than the limit influence ( $h^*$ ). As shown in Fig. 3, the validation anthocyanins of Eqs. (1)–(7) were contained within the applicability domain of each model ( $h < h^*$ ), so that their predictions could be considered reliable.

The molecular descriptors involved in the QSAR models of the current study were independent of the conformations. These descriptors only depended on the constitutional and topological features. Table 1 showed that many of these descriptors were abstract, and did not have a direct physical interpretation. Despite this, the proposed models were useful in providing a QSAR guide for the chemical synthesis of new antioxidant structures, provided that the anthocyanins, whose activity was predicted, belonged to the applicability domains of the developed models.

For example, in the case of the QSAR model given by Eq. (1) for DPPH· (17  $\mu$ M), the molecular structures of inhibitors that simultaneously had lower values of the descriptors  $GATS6p$  and  $qub1$ , would have exhibited higher predicted values of antioxidant activity  $\log(I\% + 100)$ . The reason was due to the fact that the descriptors of Eqs. (1)–(7) took positive values, and the degree of contribution of each descriptor to the predicted activity was then given by the sign and magnitude of its regression coefficient. Table 3 summarizes the simultaneous trend that the

**Table 3** Trend of the molecular descriptors for a greater predicted antioxidant activity

Equations	Descriptor contributions
(1)	$GATS6p$ (–), $qub1$ (–)
(2)	$D714$ (–), $VR1.Dzm$ (+), $VR2.Dzs$ (–)
(3)	$VR1.Dzm$ (+), $qub2$ (–), $qub3$ (–)
(4)	$qub4$ (–)
(5)	$ATSC8e$ (+), $qub5$ (+), $qub6$ (–)
(6)	$D284$ (+), $minHBint4$ (+), $qub7$ (+)
(7)	$D458$ (+)

descriptors of the QSAR Eqs. (1)–(7) should have, so that the predicted inhibition was higher for an anthocyanin structure.

Anthocyanins and anthocyanidins used in this QSAR study have shown antioxidant properties in in vitro and ex vivo tests as previously reported (Kähkönen and Heinonen 2003). Twenty-three compounds were studied. The six more common anthocyanidins were selected for this study, e.g., pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin; also, their 3-*O*-monoglycosides, 3-*O*-diglycosides, 3,5-*O*-diglycosides, 3-*O*-diglycoside-5-*O*-monoglycosides, and a 3-*O*-acyldiglycoside-5-*O*-monoglycoside were tested, including different types of hexoses in the pyranose form (see “Anthocyanins selected for this study” section). The radical scavenging activity was in vitro studied against the 1,1-diphenyl-2-picrylhydrazyl radical (DPPH·). The ability to prevent the oxidation of lipids was tested in three different lipid environments, such as emulsified and bulk methyl linoleate at different inhibitor concentrations (50 and 250  $\mu$ M, and 50  $\mu$ M, respectively), and human low-density lipoprotein (LDL) (2.5, 10, and 25  $\mu$ M). Results revealed the structural- and concentration-dependence of this activity.

The effects of structural features on the antioxidant activity have been previously analyzed according to the inhibition experimental values (Kähkönen and Heinonen 2003). As expected, hydroxylation and methoxylation patterns showed remarkable effects on the antioxidant power of the aglycones, and glycosylation/acylglycosylation pattern in the case of anthocyanins (Miguel 2011).

From the in vitro and ex vivo tests, retrospective considerations arose on the influence of substituents of the ring B and of the presence of 3-*O*-mono- or di-glycosides, 3,5-*O*-diglycosides and acyglycosides, that is, direct interpretations of the experimental values taking into account the possible mechanisms of action involved, but not always the characteristics of the medium involved. But these evaluations of the assays do not serve to design new antioxidants,

**Fig. 4** Structures of 21 anthocyanins with unknown experimental antioxidant activities

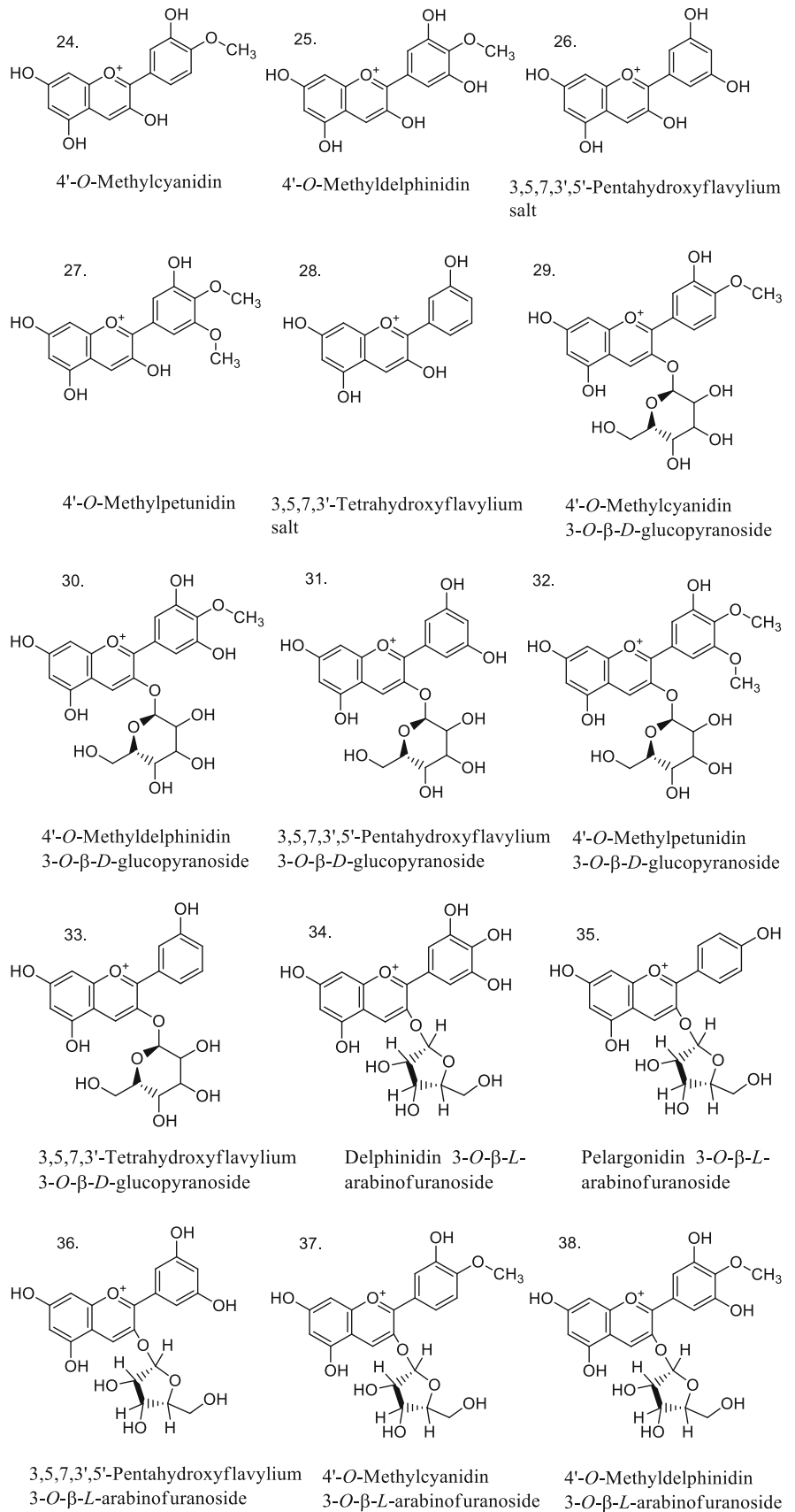
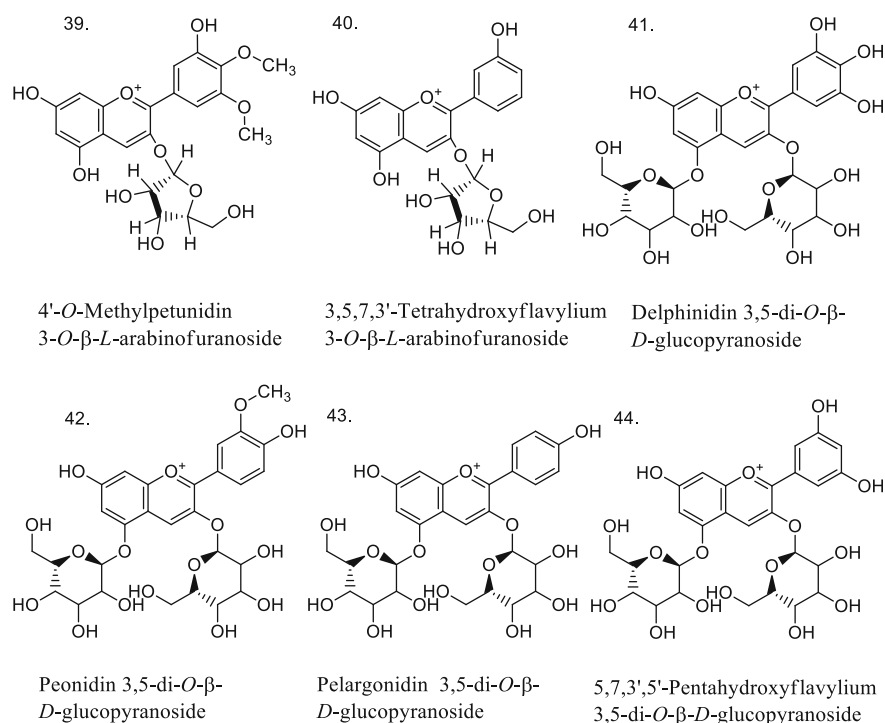


Fig. 4 continued



nor to establish a possible antioxidant value (unknown) for an anthocyanin that has not been previously tested. To be able to do this, and either to obtain numerical values of antioxidant activity of a given anthocyanin (prediction) or to design a novel antioxidant of anthocyanin nature (drug design) it was necessary to delve into the molecular electronic structure and its properties to establish the relationships with that bioactivity according to bioinformatics and statistics as currently achieved by the QSAR theory. Furthermore, statistically-based QSAR approaches are of value in Medicinal Chemistry, and help to guide lead optimization (Cherkasov et al. 2014; Mercader et al. 2016).

Therefore, we developed QSAR modeling, so that small differences at the molecular level, which are not so obvious, can be explained. For the coding process a large number of independent variables or molecular descriptors were computed and used to develop the QSAR models. Descriptors quantify several constitutional, topological, and structural properties of a molecule, but may lack the capacity for structural interpretation. Each descriptor is a reproducible and invariable value since it depends on the calculation algorithm used, while experimental bioassay values are dependent variables. A set of computational techniques were necessary for a virtual spatial design and visualization of molecules, calculation of the molecular descriptors, bioinformatics, and statistics (see “Materials and Methods” section).

The statistical predictive QSAR models obtained in the present study correlated the antioxidant bioactivity of anthocyanins with representative descriptors of their molecular structure and/or molecular properties. Results showed that different descriptors account for each antioxidant test used, and concentrations of the anthocyanins/anthocyanidins due to the test- and concentration-dependence of the antioxidant activity of these compounds.

The QSAR models of this study were obtained from the two-dimensional descriptors mentioned above. Results show that the molecular skeleton provides enough information for useful relationships, three-dimensional topography not being necessary.

Finally, a practical application of the established QSAR models of Eqs. 1–7 was performed. Antioxidant bioactivities were predicted for 21 anthocyanins/anthocyanidins with unknown experimental values (Fig. 4). As shown in Table 4, most of them can be considered as reliable predictions according to their leverage values, and most importantly, some of the proposed anthocyanins had favorable predicted bioactivities. For instance, the predicted antioxidant activity had a high value for the following compounds in different lipid environments: **26** and **38** in DPPH (17  $\mu$ M); **27**, **28**, and **39** in MeLo emulsion (50  $\mu$ M); **26**, **28**, and **31** in Human LDL (10  $\mu$ M); and **24** and **25** in MeLo in bulk (50  $\mu$ M). From the results of Table 4, experimental antioxidant tests should be carried out to verify such predicted molecular structures.

**Table 4** Predicted antioxidant activities and leverages for 21 anthocyanins with unknown experimental values

ID	Equation 1	Equation 2	Equation 3	Equation 4	Equation 5	Equation 6	Equation 7
24	5.15 (0.53)	5.00 (0.03)	5.19 (0.11)	4.71 (0.09)	4.96 (0.22)	5.45 (0.38)	5.01 (0.19)
25	5.23 (1.03)	4.73 (0.16)	5.05 (0.66)	4.66 (0.10)	4.72 (0.25)	5.36 (0.25)	5.32 (0.26)
26	5.01 (0.38)	5.00 (0.03)	5.24 (0.06)	4.70 (0.09)	5.75 (0.07)	5.03 (0.25)	4.88 (0.16)
27	4.79 (0.07)	5.04 (0.02)	5.23 (0.06)	4.74 (0.09)	4.55 (0.48)	4.21 (0.21)	4.59 (0.10)
28	4.92 (0.31)	5.04 (0.02)	5.25 (0.05)	4.70 (0.09)	5.56 (0.19)	5.35 (0.37)	4.73 (0.13)
29	4.96 (0.28)	4.84 (0.09)	5.13 (0.07)	4.46 (0.10)	4.41 (0.10)	5.23 (0.08)	4.94 (0.17)
30	4.80 (0.05)	4.92 (0.06)	5.16 (0.06)	4.47 (0.10)	4.88 (0.07)	5.33 (0.07)	4.63 (0.11)
31	4.92 (0.15)	5.02 (0.02)	5.17 (0.07)	4.55 (0.10)	5.60 (0.25)	4.99 (0.07)	4.90 (0.16)
32	4.76 (0.06)	4.48 (0.40)	5.17 (0.05)	4.46 (0.10)	4.36 (0.13)	5.10 (0.08)	4.67 (0.12)
33	4.84 (0.25)	5.03 (0.02)	5.15 (0.07)	4.48 (0.10)	4.99 (0.26)	5.19 (0.07)	4.85 (0.15)
34	4.83 (0.06)	4.87 (0.08)	5.20 (0.06)		5.19 (0.19)	5.32 (0.09)	
35	4.75 (0.06)	4.90 (0.06)	5.18 (0.07)		4.46 (0.14)	4.94 (0.09)	
36	4.86 (0.05)	4.91 (0.06)	5.20 (0.06)		5.46 (0.13)	5.08 (0.06)	
37	4.83 (0.06)	4.87 (0.08)	5.20 (0.06)		5.19 (0.19)	5.32 (0.09)	
38	5.01 (0.13)	5.01 (0.06)	5.10 (0.22)		4.60 (0.11)	5.28 (0.08)	
39	4.72 (0.20)	5.07 (0.04)	5.20 (0.04)		4.28 (0.24)	5.16 (0.15)	
40	4.79 (0.06)	4.95 (0.04)	5.19 (0.06)		4.91 (0.18)	5.30 (0.13)	
41	4.83 (0.10)	5.02 (0.03)	5.17 (0.07)		5.05 (0.27)	5.04 (0.31)	
42	4.72 (0.07)	5.02 (0.03)	5.13 (0.06)		4.17 (0.10)	5.03 (0.09)	
43	4.73 (0.06)	5.01 (0.04)	5.15 (0.07)		4.22 (0.07)	4.73 (0.12)	
44	4.83 (0.06)	5.02 (0.03)	5.16 (0.07)		5.25 (0.17)	4.86 (0.17)	

## Conclusion

Most anthocyanins and their aglycones behaved as strong antioxidants in emulsion and LDL, showing activities comparable to known antioxidants, such as  $\alpha$ -tocopherol, Trolox, catechin, and quercetin. In bulk methyl linoleate, anthocyanins and anthocyanidins showed only weak antioxidant or even oxidation-promoting activities. Mostly, the activities of the glycosides and aglycones did not differ markedly in emulsion. In LDL, the aglycones generally showed higher activities than glycosides. In contrast, in bulk oil, glycosides were more effective than aglycones.

The QSAR models obtained in this work allowed a prospective analysis leading to proper antioxidant prediction and design of new antioxidant structures in the anthocyanin field, depending solely on constitutional and electrotopological features of the molecular structures. Therefore, these results allowed to establish quantitative structure–activity relationships that are satisfactory for predicting properties through molecular descriptors; and to show that QSAR studies are suitable for designing molecules with optimal desired antioxidant activities.

**Acknowledgements** We thank the National Scientific and Technical Research Council of Argentina [Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina] (PIP0311) and Universidad de Buenos Aires (Argentina) for financial support; and Secretaría de Ciencia, Tecnología e Innovación Productiva (formerly Ministerio de Ciencia, Tecnología e Innovación Productiva) for

electronic library facilities. N.A.S. thanks the Scientific Research Commission [Comisión de Investigaciones Científicas (CIC), La Plata city, Argentina] for a fellowship. A.B.P. and P.R.D. are Research Members of CONICET.

## Compliance with ethical standards

**Conflict of interest** All authors declare that they have no conflict of interest.

## References

- ACD/ChemSketch (2016) [www.acdlabs.com](http://www.acdlabs.com)
- Bentz EN, Pomilio AB, Lobayan RM (2017) Donor–acceptor interactions as descriptors of the free radical scavenging ability of flavans and catechin. *Comput Theor Chem* 1110:14–24
- Bonesi M, Leporini M, Tenuta MC, Tundis R (2019) The role of anthocyanins in drug discovery: recent developments. *Curr Drug Discov Technol.* <https://doi.org/10.2174/1570163816666190125152931>
- Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: Where have you been? Where are you going to? *J Med Chem* 57:4977–5010
- 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. *J Chem Inf Model* 52:2044–2058

- de Sousa Moraes LF, Sun X, Peluzio MDCG, Zhu MJ (2019) Anthocyanins/anthocyanidins and colorectal cancer: What is behind the scenes? *Crit Rev Food Sci Nutr* 59:59–71
- Du H, Lai L, Wang F, Sun W, Zhang L, Li X, Wang L, Jiang L, Zheng Y (2018) Characterisation of flower colouration in 30 *Rhododendron* species via anthocyanin and flavonol identification and quantitative traits. *Plant Biol (Stuttg)* 20:121–129
- Duchowicz PR (2018) Linear regression QSAR models for Polo-Like Kinase-1 Inhibitors. *Cells* 7:1–11
- Duchowicz PR, Castro EA, Fernández FM (2006) Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *MATCH Commun Math Comput Chem* 55:179–192
- Epi Suite 4.11 (2018) <https://www.epa.gov/tsca-screening-tools/epi-suite-estimation-program-interface>
- Fujiwara Y, Kono M, Ito A, Ito M (2018) Anthocyanins in *Perilla* plants and dried leaves. *Phytochemistry* 147:158–166
- Harborne JB, Williams CA (2000) Advances in flavonoid research since 1992. *Phytochemistry* 55:481–504
- He J, Giusti MM (2010) Anthocyanins: natural colorants with health-promoting properties. *Annu Rev Food Sci Technol* 1:163–187
- Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L, Su Z, Perkin R, Tong W (2008) Mold(2), molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J Chem Inf Model* 48:1337–1344
- Jaganath IB, Crozier A (2010) Dietary flavonoids and phenolic compounds. In: Fraga CG (ed) *Plant phenolics and human health: biochemistry, nutrition, and pharmacology*. Wiley, Hoboken
- Jiang X, Li X, Zhu C, Sun J, Tian L, Chen W, Bai W (2019) The target cells of anthocyanins in metabolic syndrome. *Crit Rev Food Sci Nutr* 59:921–946
- Kähkönen MP, Heinonen M (2003) Antioxidant activity of anthocyanins and their aglycons. *J Agric Food Chem* 51:628–633
- Kaurinovic B, Vastag D (2019) Flavonoids and phenolic acids as potential natural antioxidants. *intechopen*. Open access peer-reviewed chapter—online first. <https://doi.org/10.5772/intechopen.83731>. <https://www.intechopen.com/online-first/flavonoids-and-phenolic-acids-as-potential-natural-antioxidants>. Accessed 10 June 2019
- Khan MS, Ali T, Kim MW, Jo MH, Chung JI, Kim MO (2019) Anthocyanins improve hippocampus-dependent memory function and prevent neurodegeneration via JNK/Akt/GSK3 $\beta$  signaling in LPS-treated adult mice. *Mol Neurobiol* 56:671–687
- Krga I, Milenkovic D (2019) Anthocyanins: from sources and bioavailability to cardiovascular-health benefits and molecular mechanisms of action. *J Agric Food Chem* 67:1771–1783
- Lavine BK, Davidson CE, Breneman C, Katt W, Sundling CM (2003) Electronic van der Waals surface property descriptors and genetic algorithms for developing structure-activity correlations in olfactory databases. *J Chem Inf Comput Sci* 43:1890–1905
- Li S, Wu B, Fu W, Reddivari L (2019) The anti-inflammatory effects of dietary anthocyanins against ulcerative colitis. *Int J Mol Sci*. <https://doi.org/10.3390/ijms20102588>
- Ma X, Ning S (2019) Cyanidin-3-glucoside attenuates the angiogenesis of breast cancer via inhibiting STAT3/VEGF pathway. *Phytother Res* 33:81–89
- Matlab 7.6 <http://www.mathworks.com>
- Mercader AG, Duchowicz PR, Fernández FM, Castro EA (2010) Replacement method and enhanced replacement method *versus* the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR Theories. *J Chem Inf Model* 50:1542–1548
- Mercader AG, Duchowicz PR, Sivakumar PM (eds) (2016) *Chemo-metrics applications and research: QSAR in medicinal chemistry*. CRC Press, Boca Raton
- Miguel MG (2011) Anthocyanins: antioxidant and/or anti-inflammatory activities. *J Appl Pharm Sci* 1:7–15
- PaDEL (2016). <http://www.yapcsoft.com/dd/padeldescriptor>
- Pomilio AB, Mercader AG (2017) Natural acylated anthocyanins and other related flavonoids: structure elucidation of *Ipomoea cairica* compounds and QSAR studies including multidrug resistance. In: Atta-ur-Rahman D (ed) *Studies in natural products chemistry. (Bioactive natural products)*. Elsevier, The Netherlands, pp 291–321
- Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) Some case studies on application of “r2 m” metrics for judging quality of Quantitative Structure-Activity Relationship predictions: emphasis on scaling of response data. *J Comput Chem* 34:1071–1082
- Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Syst* 145:22–29
- Sakuta M (2014) Diversity in plant red pigments: anthocyanins and betacyanins. *Plant Biotechnol Rep* 8:37–48
- Sousa A, Araújo P, Azevedo J, Cruz L, Fernandes I, Mateus N, de Freitas V (2016) Antioxidant and antiproliferative properties of 3-deoxyanthocyanidins. *Food Chem* 192:142–148
- Thankam Finosh G, Jayabalan M (2013) Reactive oxygen species—control and management using amphiphilic biosynthetic hydrogels for cardiac applications. *Adv Biosci Biotechnol* 4:1134–1146
- Tsakiroglou P, VandenAkker NE, Del Bo’ C, Riso P, Klimis-Zacas D (2019) Role of berry anthocyanins and phenolic acids on cell migration and angiogenesis: an updated overview. *Nutrients*. <https://doi.org/10.3390/nu11051075>
- Ullah R, Khan M, Shah SA, Saeed K, Kim MO (2019) Natural antioxidant anthocyanins—a hidden therapeutic candidate in metabolic disorders with major focus in neurodegeneration. *Nutrients*. <https://doi.org/10.3390/nu11061195>
- Valdes-Martini JR, García Jacas CR, Marrero-Ponce Y, Silveira Vaz’d Almeida Y, Morrel C (2012) Versión 1.0. CAMD-BIR Unit, CENDA Number of register: 2373-2012
- Vishnu VR, Renjith RS, Mukherjee A, Anil SR, Sreekumar J, Jyothi AN (2019) Comparative study on the chemical structure and in vitro antiproliferative activity of anthocyanins in purple root tubers and leaves of sweet potato (*Ipomoea batatas*). *J Agric Food Chem* 67:2467–2475
- Vitale AA, Bernatene EA, Vitale MG, Pomilio AB (2016) New insights of the Fenton reaction using glycerol as experimental model. Effect of O $_2$ , inhibition by Mg $^{2+}$ , and oxidation state of Fe. *J Phys Chem A* 120:5435–5445
- Wongwachai T, Teeyakasem P, Pruksakorn D, Kongtawelert P, Pothacharoen P (2019) Anthocyanins and metabolites from purple rice inhibit IL-1 $\beta$ -induced matrix metalloproteinases expression in human articular chondrocytes through the NF- $\kappa$ B and ERK/MAPK pathway. *Biomed Pharmacother* 112:108610
- Zhang ZC, Zhou Q, Yang Y, Wang Y, Zhang JL (2019) Highly acylated anthocyanins from purple sweet potato (*Ipomoea batatas* L.) alleviate hyperuricemia and kidney inflammation in hyperuricemic mice: possible attenuation effects on allopurinol. *J Agric Food Chem* 67:6202–6211

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.