Article

J. Braz. Chem. Soc., Vol. 23, No. 9, 1747-1755, 2012.
Printed in Brazil - ©2012 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

A Multivariate Statistical Analysis of the Quantitative Structure-Activity Relationships (QSAR) of 2-(Substituted Phenyl)indan-1,3-diones with Hypolipidemic Activity

Aluizio G. da Silva*[*,a] and Mozart N. Ramos[b]

[a]*Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, 55590-000 Ipojuca-PE, Brazil*

[b]*Departamento de Química Fundamental, Universidade Federal de Pernambuco, 50740-540 Recife-PE, Brazil*

Descritores eletrônicos obtidos de cálculos de orbitais moleculares AM1, juntamente com os parâmetros log P e π, foram empregados para obter relações quantitativas estrutura-atividade (QSAR) para um conjunto de dezesseis derivados de 2-fenil-indan-1,3-dionas possuindo atividade hipolipidêmica. Os descritores AM1 foram obtidos tanto para o modelo da molécula isolada como também para o modelo COSMO (modelo de seleção considerando condutor ideal) visando incluir o efeito do solvente em nossa análise QSAR. Dos dezesseis derivados estudados, três deles foram excluídos do conjunto de treinamento tendo por objetivo servir como conjunto de teste. Em geral, os resultados obtidos deste conjunto de teste validaram nossos modelos quantitativos, os quais foram baseados em análise de regressão linear e de componentes principais (PCA). A partir disso, novos derivados foram previstos para ser agentes hipolipidêmicos mais potentes do que os derivados de 2-fenil-indan-1,3-dionas cujas atividades experimentais foram determinadas. Por exemplo, nossos resultados sugerem que 2-*orto*-hidroxi-fenil-indan-1,3-diona é um potencial candidato para novas triagens de estudos experimentais na redução dos níveis de colesterol e de triglicerídeos.

Electronic descriptors obtained from AM1 molecular orbital calculations were employed, along with the empirical parameters log P and π, in a multivariate statistical analysis of the quantitative structure-activity relationships (QSAR) in a group of sixteen 2-(substituted phenyl)indan-1,3-diones possessing hypolipidemic activities. AM1 descriptors were obtained both for the isolated molecule model and, to simulate the solvent effect, for the COSMO model (conductor-like screening model) in our QSAR analysis. From sixteen studied derivatives, three were excluded from the set of molecules used in the model building stage in order to serve as a test set. In general, the results obtained from this test set validated our quantitative models, which were based on linear regressions and principal component analysis (PCA). From these results, new derivatives were predicted to be more potent hypolipidemic agents than the indan-1,3-diones for which experimental activities were determined. For instance, our results suggest that 2-(*ortho*-hydroxy-phenyl)indan-1,3-dione is a potential candidate for new experimental screening studies for reducing serum cholesterol and triglyceride levels.

Keywords: hypolipidemic activity, multivariate statistical analysis, AM1 calculations

Introduction

Cyclic imides are an important family of organic compounds with therapeutic potential. Cechinel Filho *et al.*[1] presented an interesting review on the chemical and biological aspects of several sub-classes of this family, such as maleimides, phthalimides and succinimides. From the point of view of toxicity of these compounds, few studies are found in current literature. One of these studies, performed by Hall's group from North Carolina at Chapel Hill,[2] showed that phthalimide derivatives have no significant toxicity towards rodents, as well as in their fertility.

Recently, El-Zahabi *et al.*[3] synthesized new nitrogen-substituted derivatives of cyclic imides: phthalimide,

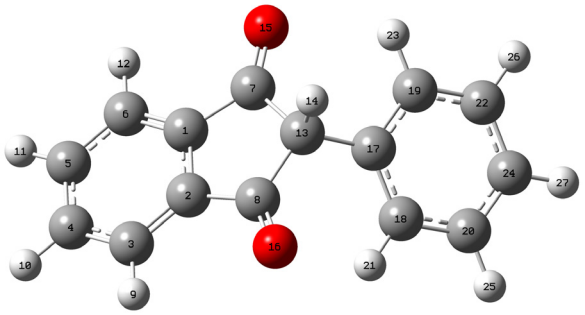
*e-mail: aluizioufpe@hotmail.com

1,8-naphtalimide and diphenimide. A preliminary evaluation of the hypolipidemic activity of these newly prepared compounds showed that several derivatives significantly decrease serum total cholesterol and triglyceride levels in rats with triton WR-1339-induced hyperlipidemia. The dose used (150 mg kg^{-1} via intraperitoneal) was comparable to Fenofibrate, which is one of the second generation of fibrate drugs.

Some studies investigated the hypolipidemic activity of a set of phthalimide derivatives injected intraperitoneally in CF_1 male mice.[4] Most of the studied compounds showed potent activity in reducing serum cholesterol and triglyceride levels at the optimum dose of 20 mg kg^{-1} *per* day.[4] Phthalimide and its derivatives are effective in reducing liver mitochondrial citrate exchange, acetyl-CoA synthetase, acetyl-CoA carboxylase and phosphatidate phosphohydrolase activities, as well as the levels of liver and small intestine lipids at 20 mg kg^{-1} *per* day. For instance, phthalimide administration to rats was related to elevated excretion of cholesterol in the bile, with reduction of cholesterol and triglyceride contents in the blood lipoprotein fractions, but with an increase in phospholipid content. A sixteen-day treatment with *N*-phenylphthalimide, for example, reduced the serum cholesterol level to 57% of its initial value. The substitution on the phenyl ring drastically affected hypolipidemic activity. When an ethyl group was introduced, the *N*-phenylphthalimide lost its effect. On the other hand, the substitution of an ethoxy group in the *ortho* position increased the hypolipidemic activity, and the cholesterol level fell to 43% after the treatment.

A few years ago,[5] our research group obtained electronic descriptors from AM1 molecular orbital calculations, as well as the empirical descriptors log P and π, in order to better understand the quantitative structure-activity relationships (QSAR) for several phthalimide derivatives.[6,7] For the derivatives without the phenyl *N*-substitution,[6] our group found that the LUMO orbital energy and the carbonyl group polarity are important parameters for explaining the dependence of hypolipidemic activity on molecular structure. Likewise, the group successfully explained this dependence in *N*-phenylphthalimide derivatives.[7] Quantitative models based on multiple linear regression and on partial least squares regression allowed us to suggest new derivatives that are expected to be more potent hypolipidemic agents than those employed in the model building stage. For example, the *ortho*-NO-phenylphthalimide was expected to be the most effective of the new compounds.

Murthy *et al.*[8] also investigated the hypolipidemic activity in 2-(substituted phenyl)indan-1,3-diones. It is interesting to point out that the synthetic methods to prepare these derivatives are largely reported in the literature.[9] For example, 2-methylindan-1,3-dione was prepared according to the procedure of Mosher and Soeder[10] by the condensation of dimethyl phthalate with 3-pentanone in the presence of sodium hydride. All the substituted phenyl analogues were synthesized following the method of Freedman *et al.*[11] Chapman *et al.*[4] reported that the parent compound, 2-phenylindan-1,3-dione (Scheme 1), is significantly active in reducing cholesterol and triglyceride levels by 40% after 16 days of administration in CF_1 male mice at 20 mg kg^{-1} *per* day via intraperitoneal. They also studied the influence of the phenyl ring substitution on the hypolipidemic activity of these compounds. The substitution of the phenyl ring provided mixed results. For example, changing the methoxy group at the *para* position in the phenyl ring to ethoxy maintained hypocholesterolemic activity but lowered hypotriglyceridemic activity markedly from 58 to 19%. The substitution of methyl groups in the *ortho* and *meta* positions of the phenyl ring gave rise to compounds that clearly possessed lower hypocholesterolemic activity in mice than 2-phenylindan-1,3-dione, in contrast to what occurs when the methyl group is in the *para* position. On the other hand, the introduction of methyl groups in the *ortho*, *meta* and *para* positions of the phenyl ring increased hypotriglyceridemic activity. They also observed that 2-(4-methoxyphenyl)indan-1,3-dione was one of the more active compounds with 41% reduction of serum cholesterol and 58% reduction of serum triglyceride levels on day 16. A more careful look on the effects of this derivative demonstrated that key enzymes in the *de novo* synthesis of lipids were inhibited by the drug. The levels of lipids in tissues decreased, but those in the feces were raised. These alterations in lipid content on rat lipoprotein fractions by the drug appeared favorable.



Scheme 1. Structure of 2-phenylindan-1,3-dione (**I**).

In an attempt to discover which structural modifications contribute to these pharmacological changes, it was examined a set of sixteen 2-(substituted phenyl)indan-1,3-diones, in search of a relation between the experimentally

determined hypolipidemic activities and electronic parameters obtained from AM1 molecular orbital calculations. These calculations were performed with or without considering the solvent effect. In this study, it was also included the log P and π empirical parameters.

Methodology

To obtain the electronic properties of the compounds considered in this work, quantum-chemical calculations were performed at AM1 semi-empirical level using the Gaussian program.[12] The molecular geometry of each compound was fully optimized and no imaginary frequency was observed. From the wave-functions thus obtained, it was extracted the frontier orbital energies, Mulliken net atomic charges and molecular dipole moments. These variables, together with the log P (partition coefficient) and π (Hansch's parameter) empirical parameters, form a set of 12 descriptors for each 2-(substituted phenyl)indan-1,3-dione. The octanol-water partition coefficient values were estimated using the Kowwin version 1.67 program.[13]

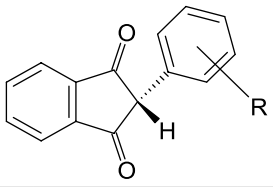
Hypolipidemic activity is expressed in terms of pC_{16} and pT_{16}, which are the logarithms of the reduction in cholesterol and triglyceride levels observed after sixteen-day period of administering the drug in CF_1 male mice, respectively. Thus, the higher the pC_{16} and pT_{16} values, the more active the compound is in terms of the reduction in cholesterol and triglyceride levels, respectively. Table 1 shows the sixteen 2-(substituted phenyl)indan-1,3-diones studied in this work with the corresponding figures for reduction in cholesterol and triglyceride levels observed in CF_1 male mice after the drug administration period. From those sixteen compounds, three were excluded from the training set to serve as a test set for validation of the mathematical models here obtained.

Quantitative structure-activity relationship (QSAR) models relating these activity values to the molecular descriptors were investigated by means of two multivariate techniques: traditional multiple linear regression (MLR) and principal component analysis (PCA).[14] The latter was used because some of the twelve electronic descriptors have a high degree of inter-correlation, which could cause statistical instability in the ordinary multiple regression equations, as a result of multicollinearity. Therefore, it was first performed a principal component analysis on them, and regressed the dependent variable, hypolipidemic activity, on the scores obtained from the PCA, rather than on the descriptors themselves.

In a principal component analysis, the original data matrix is projected onto a subspace defined by linear combinations of the original variables with maximum variance (that is, maximum information) and orthogonal to each other.[14] The first principal component (PC1) is the axis describing the maximum possible variance in the original multidimensional space; the second component is orthogonal to PC1 and describes the maximum variance not already described by PC1, and so on. The orthogonality between different PC axes eliminates the regression multicollinearity problem. Each PC is characterized by three mathematical entities: (*i*) the percentage of explained variance; (*ii*) a loading vector, whose elements are the cosines of the angles that the PC axis forms with the original variable axes; and (*iii*) a scores vector, containing the coordinates locating the individual compounds on the PC axis. A regression on the scores of a PCA is called a principal component regression.[15]

In order to obtain the electronic descriptors simulating the solvent effect, it was used the conductor-like screening model (COSMO)[16] in MOPAC 09. This dielectric continuum model[17] was used here to simulate the solvation phenomenon using water as a solvent.

Table 1. Hypolipidemic activities of the sixteen 2-(substituted phenyl)indan-1,3-diones studied in this work



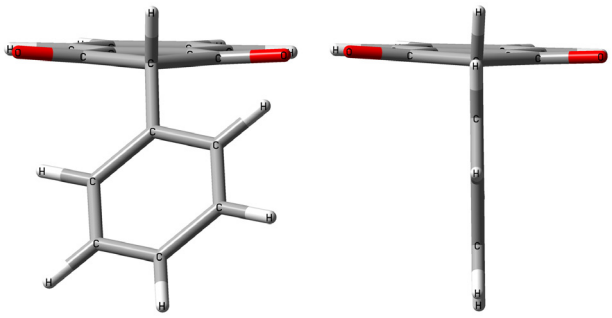
Training set	($\%C_{16}$)	($\%T_{16}$)
H	40 ± 2	40 ± 4
o-CH_3	31 ± 3	51 ± 1
m-CH_3	29 ± 3	49 ± 4
p-CH_3	35 ± 3	58 ± 2
o-Cl	44 ± 4	45 ± 3
m-Cl	40 ± 4	30 ± 3
p-Cl	37 ± 3	53 ± 2
o-COOH	22 ± 2	48 ± 4
p-OCH_2CH_3	41 ± 2	19 ± 4
p-COOH	22 ± 2	5 ± 2
o-OCH_3	28 ± 2	52 ± 3
m-OCH_3	33 ± 5	45 ± 4
p-OCH_3	41 ± 3	58 ± 4
2,4-Dichloro[a]	27 ± 6	13 ± 5
3,4-Dimethoxy[a]	31 ± 3	34 ± 8
3,4-Dichloro[a]	16 ± 5	11 ± 4

[a]Molecules used as the test set in this work.

Results and Discussion

The conformational analysis performed in the parent compound (**I**, Scheme 1) reveals that the $C_8C_{13}C_{17}C_{18}$

dihedral angle is 7.5° for its most stable structure (Figure 1a), corresponding to a gauche conformation. In the maximum energy structure that same dihedral angle is 52.5° for an orthogonal conformation (Figure 1b).



(a) (b)

Figure 1. Structures of the parent compound (**I**) in the most stable conformation (a) and in the maximum energy conformation (b).

In order to test these AM1 values, it was performed B3LYP/6-31G(d,p) molecular orbital calculations.[18] The $C_8C_{13}C_{17}C_{18}$ dihedral angles obtained from these calculations for the minimum and maximum energy structures were 7.1° and 57.1°, respectively. Both, therefore, agree very well with the AM1 angles.

For the 2-(substituted phenyl)indan-1,3-diones with substituents in the phenyl ring *meta* and *para* positions, our results reveal a very similar conformation to that in the parent compound, with the $C_8C_{13}C_{17}C_{18}$ dihedral angle only varying by ± 1°. On the other hand, this angle varies substantially when the substituent is located in the *ortho* position, as observed with the *N*-phenylphthalimides.[4] The actual value depends on both the electronic nature and the size of this *ortho* substituent. Thus, the $C_8C_{13}C_{17}C_{18}$ dihedral angles for the $–OCH_3$, $–CH_3$, –Cl and –COOH substituents in the *ortho* position are: 26.8°, 42.5°, 33.8° and 59.4°, respectively. Therefore, their minimum energy conformations correspond to orthogonal structures and are thus similar to those found at the maximum points in the *meta* and *para* substituents. The maximum energy points for these substituents in the *ortho* position occur near 150°. As a consequence, the potential surface profile of the latter is very different from that found in the parent compound and for the *meta* and *para* substituents.

Table 2 shows the figures for the most significant descriptors, which have the highest correlation coefficients with pC_{16} and pT_{16}. The inclusion of the solvent effect in the AM1 calculations using the COSMO method does not significantly alter such descriptors. The latter are given in parentheses.

Table 2. Hypolipidemic activities and most significant descriptors for the thirteen 2-(substituted phenyl)indan-1,3-diones studied in this work. Values of the electronic descriptors obtained from AM1/COSMO calculations are given in parentheses

Compound	–R	pC_{16}	pT_{16}	ε_{LUMO} (eV)	μ (D)	qH_{14} (e^-)	qC_{17} (e^-)	π	log P
1	H	1.602	1.602	–0.981 (–1.037)	2.496 (3.417)	0.153 (0.183)	–0.072 (–0.077)	0.000	2.970
2	*o*-CH_3	1.491	1.708	–0.952 (–1.028)	2.302 (3.220)	0.152 (0.181)	–0.067 (–0.083)	0.560	3.300
3	*m*-CH_3	1.462	1.690	–0.965 (–1.036)	2.323 (3.225)	0.153 (0.183)	–0.068 (–0.072)	0.560	3.360
4	*p*-CH_3	1.544	1.763	–0.962 (–1.035)	2.372 (3.285)	0.152 (0.182)	–0.078 (–0.084)	0.560	3.360
5	*o*-Cl	1.643	1.653	–0.907 (–1.057)	1.327 (4.410)	0.148 (0.182)	–0.048 (–0.069)	0.710	3.520
6	*m*-Cl	1.602	1.477	–1.067 (–1.059)	3.453 (4.634)	0.155 (0.187)	–0.061 (–0.062)	0.710	3.580
7	*p*-Cl	1.568	1.724	–1.077 (–1.056)	3.490 (4.648)	0.155 (0.187)	–0.070 (–0.068)	0.710	3.580
8	*o*-COOH	1.342	1.681	–1.079 (–1.075)	3.310 (4.532)	0.170 (0.195)	–0.013 (–0.012)	–0.320	2.480
9	*p*-OCH_2CH_3	1.613	1.279	–0.959 (–1.035)	2.877 (4.133)	0.151 (0.181)	–0.111 (–0.112)	0.380	3.320
10	*p*-COOH	1.342	0.699	–1.156 (–1.070)	4.673 (6.079)	0.158 (0.191)	–0.032 (–0.027)	–0.320	2.630
11	*o*-OCH_3	1.447	1.716	–0.843 (–1.019)	1.003 (1.825)	0.169 (0.185)	–0.086 (–0.113)	–0.020	2.840
12	*m*-OCH_3	1.519	1.653	–0.956 (–1.044)	2.408 (3.420)	0.155 (0.186)	–0.039 (–0.041)	–0.020	2.900
13	*p*-OCH_3	1.613	1.763	–0.970 (–1.035)	2.981 (4.187)	0.151 (0.182)	–0.109 (–0.111)	–0.020	2.910

Table 3. Matrix of the correlation coefficients of the experimental hypolipidemic activities and most significant descriptors. Correlation coefficients for the AM1/COSMO electronic descriptors are given in parentheses

	pC_{16}	pT_{16}	ε_{LUMO}	μ	qH_{14}	qC_{17}	π	log P
pC_{16}	1.00 (1.00)	0.34 (0.34)	0.38 (0.32)	−0.31 (−0.11)	−0.73 (−0.71)	−0.60 (−0.58)	0.60 (0.60)	0.69 (0.69)
pT_{16}	0.34 (0.34)	1.00 (1.00)	0.58 (0.43)	−0.64 (−0.66)	−0.02 (−0.34)	−0.22 (−0.33)	0.35 (0.35)	0.27 (0.27)
ε_{LUMO}	0.38 (0.32)	0.58 (0.43)	1.00 (1.00)	−0.95 (−0.84)	−0.13 (−0.78)	−0.49 (−0.84)	0.21 (0.19)	0.19 (0.19)
μ	−0.31 (−0.11)	−0.64 (−0.66)	−0.95 (−0.84)	1.00 (1.00)	0.03 (0.49)	0.27 (0.55)	−0.27 (−0.11)	−0.21 (−0.06)
qH_{14}	−0.73 (−0.71)	−0.02 (−0.34)	−0.13 (−0.78)	0.03 (0.49)	1.00 (1.00)	0.40 (0.81)	−0.60 (−0.55)	−0.66 (−0.58)
qC_{17}	−0.60 (−0.58)	−0.22 (−0.33)	−0.49 (−0.84)	0.27 (0.55)	0.40 (0.81)	1.00 (1.00)	−0.30 (−0.34)	−0.37 (−0.38)
π	0.60 (0.60)	0.35 (0.35)	0.21 (0.19)	−0.27 (−0.11)	−0.60 (−0.55)	−0.30 (−0.34)	1.00 (1.00)	0.98 (0.98)
log P	0.69 (0.69)	0.27 (0.27)	0.19 (0.19)	−0.21 (−0.06)	−0.66 (−0.58)	−0.37 (−0.38)	0.98 (0.98)	1.00 (1.00)

Table 3 shows the matrix of correlation coefficients for both situations, i.e., isolated and solvated compounds of 2-(substituted phenyl)indan-1,3-diones. This matrix shows that there is no linear correlation between hypocholesteremic and hypotriglyceridemic activity, as the correlation coefficient is only 0.34. The activities are therefore treated separately.

Hypocholesteremic activity

Table 3 shows that the hydrogen atomic charge bonded to the sp^3 carbon of the indan group (Scheme 1) is the better descriptor for the structure- hypocholesteremic activity relationship in 2-(substituted phenyl)indan-1,3-diones. The correlation coefficients are −0.73 and −0.71 for the isolated and solvated compounds, respectively. This implies that compounds with a less positive hydrogen charge (qH_{14}) will tend to be more hypocholesteremic. The regression equation is shown to follow the isolated molecule model:

$$pC_{16} = 2.24(\pm 0.48) - 11.02(\pm 3.11)\ qH_{14} \quad (1)$$

with n = 13, R = – 0.73, R^2 = 0.53, F = 12.57, s = 0.072 and p = 0.0046. The values in parentheses are the standard error estimates of the regression coefficients. Though the R^2 value is not very impressive, both coefficients are statistically significant at the 95% confidence level. Figure 2 plots the experimental pC_{16} values against those predicted using equation 1. It can be seen that greater linear deviations are found for the *p*-COOH and *o*-OCH_3 substituents.

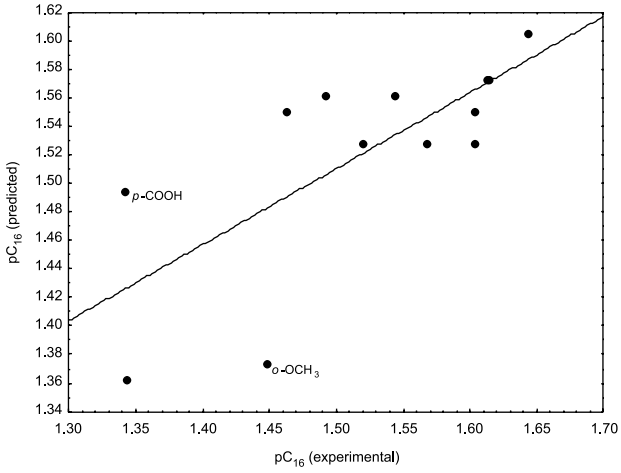


Figure 2. Plot of the AM1 predicted hypocholesteremic activities using equation 1 against their corresponding experimental values.

Since the hydrogen charge is influenced by the conformational effect, "optimum" values for this descriptor were analyzed as a function of the $C_8C_{13}C_{17}C_{18}$ dihedral angle for both substituents. Conformational analysis reveals that these optimum values are 0.159 and 0.160 e^- for the *p*-COOH and *o*-OCH_3 substituents, respectively. These values were used to obtain the following regression equation:

$$pC_{16} = 4.71(\pm 0.67) - 20.66(\pm 4.37)\ qH_{14} \quad (2)$$

with n = 13, R = – 0.82, R^2 = 0.67, F = 22.38, s = 0.060 and p = 0.00062. This equation can explain 67% of the variance in activity at a 95% confidence level. Figure 3 shows the predicted pC_{16} activities using equation 2 compared with

the experimental pC_{16} values. As expected, there is a better linear fit.

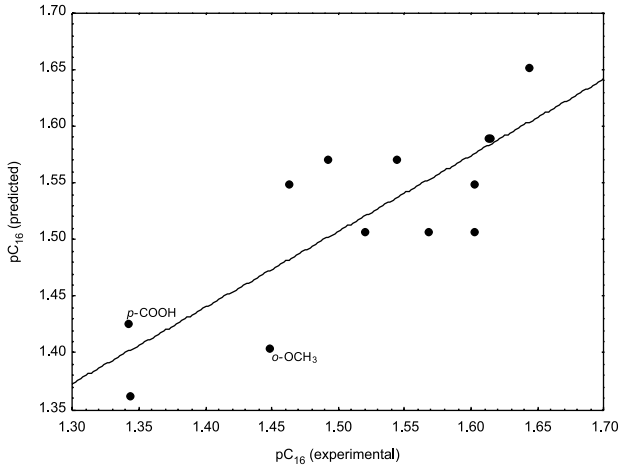


Figure 3. Plot of the AM1 predicted hypocholesteremic activities using equation 2 against their corresponding experimental values.

Neto *et al.*[19] remarked that the examination of the residuals left by mathematical models is as important as the *F*-test. Figure 4 shows that the residuals are randomly distributed. This stresses the existence of a linear relationship between the hypocholesteremic activity and the hydrogen atomic charge bonded to the sp^3 carbon of the indan group, as shown in equation 2.

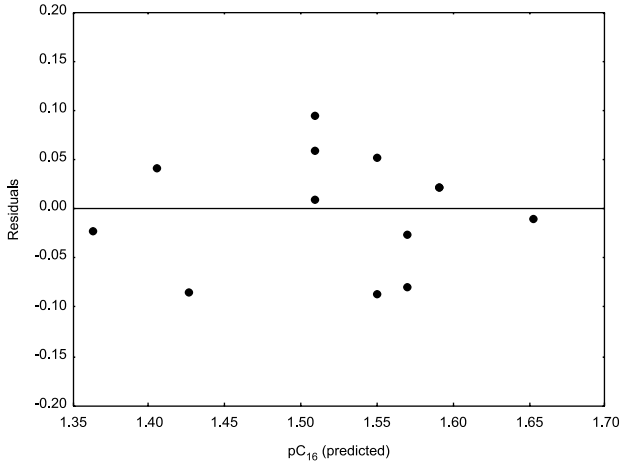


Figure 4. Plot the residuals *vs.* predicted values obtained from equation 2.

Table 4 shows the percentage reduction in hypocholesteremic activity levels obtained from equation 2 and from experimental results. In general, it is possible to verify that there is a very good agreement between the predicted and the experimental activities in considering both the training and test sets, except for 2-(3,4-dichlorophenyl)indan-1,3-dione.

For the solvated molecules, the regression equation is similar to equation 1 obtained for the isolated molecules

Table 4. Reduction percents of hypocholesteremic activity levels obtained from equation 2 and from experimental results

Training set	$(\%C_{16})_{exp}$	$(\%C_{16})_{pred}$
H	40 ± 2	35
o-CH_3	31 ± 3	37
m-CH_3	29 ± 3	35
p-CH_3	35 ± 3	37
o-Cl	44 ± 4	45
m-Cl	40 ± 4	32
p-Cl	37 ± 3	32
o-COOH	22 ± 2	23
p-OCH_2CH_3	41 ± 2	39
p-COOH	22 ± 2	27
o-OCH_3	28 ± 2	25
m-OCH_3	33±5	32
p-OCH_3	41 ± 3	35
Test set	$(\%C_{16})_{exp}$	$(\%C_{16})_{pred}$
2,4-Dichloro	27 ± 6	21
3,4-Dimethoxy	31 ± 3	29
3,4-Dichloro	16 ± 5	34

using the most stable conformation for each compound, as seen from the following equation:

$$pC_{16} = 4.68(\pm 0.94) - 17.04(\pm 5.07)\ qH_{14} \quad (3)$$

with n = 13, R = – 0.71, $R^2 = 0.51$, F = 11.28, s = 0.074 and $p = 0.0064$. Again, the values in parentheses are the standard error estimates for the regression coefficients. Though the R^2 value is not very impressive, both coefficients are statistically significant at a 95% confidence level.

Principal component analysis (PCA)

Table 3 shows that qC_{17}, log P and π are also reasonably well correlated with pC_{16}. However, they are also reasonably well correlated with qH_{14}. These descriptors could thus cause statistical instability in the ordinary multiple regression equation as a result of multicollinearity. In order to resolve this question, a principal component analysis was performed, and the dependent variable (pC_{16}) regressed on the scores obtained from the PCA, rather than the descriptors themselves. A PCA based on the correlation matrix of the four descriptors (qH_{14}, qC_{17}, log P and π) was carried out, using the AM1 values obtained from the most stable conformations for the isolated molecules. The loading plot for the first two components (PC1 and PC2), which together account for 90% of the total information,

is shown in Figure 5a. PC1 retains 69% of this information whereas PC2 retains 21%. Hypocholesteremic activity is also projected on this plot, but was not used in the PCA calculations.

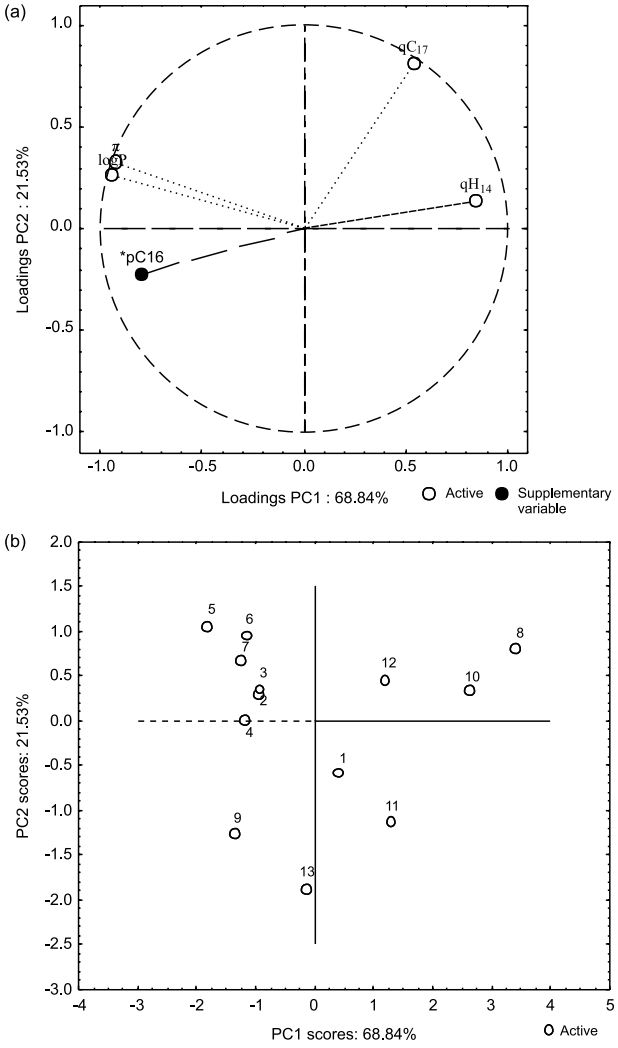


Figure 5. Results of the principal component analysis performed on the qH_{14}, qC_{17}, log P and π descriptors: (a) loading plot and (b) score plot. The percent information reproduced by each component is given in the respective axis title.

This plot shows that hypocholesteremic activity is on the same side as the log P and π descriptors with negative loadings, whereas qH_{14} and qC_{17} show positive loadings in PC1. This means that the compounds that are more active in reducing cholesterol levels are expected to be associated with high log P and π values and with low qH_{14} and qC_{17} values. It is worth noting that qH_{14} is the most important descriptor for the hypocholesteremic activity since it practically coincides with the positive side of the PC1 axis and is the exact opposite of pC_{16}. The corresponding score plot is shown in Figure 5b.

Overall, the more active compounds have more negative scores on PC1. The least active compound **8** with the *o*-COOH substituent has the most positive PC1 score, whereas the most active is compound **5** with the *o*-Cl substituent, which has the most negative PC1 score. The score in the first component can thus be interpreted as a composite theoretical measure of hypocholesteremic activity.

A regression of the pC_{16} values on the PC1 scores yields equation 4:

$$pC_{16} = 1.52(\pm 0.02) - 0.04(\pm 0.01)\, t_1 \quad (4)$$

with n = 13, R = –0.78, $R^2 = 0.61$, F = 17.44, s = 0.065 and $p = 0.0016$, where t_1 stands for the score on the PC1 axis. This equation can explain 61% of the activity variance at a 95% confidence level. Table 5 shows the percentage reduction in hypocholesteremic activity levels obtained from equation 4 and the experimental values. In general, it is possible to verify that there is a very good agreement between them, especially for the molecules employed for the training set. Again, it was verified a large discrepancy between the values for the 2-(3,4-dichlorophenyl)indan-1,3-dione used in the testing set. Furthermore, it is important to point out that the agreement to the 2-(3,4-dichlorophenyl)indan-1,3-dione is not as good as that verified above using the equation 2.

Table 5. Reduction percents of hypocholesteremic activity levels obtained from equation 4 and from experimental results

Training set	$(\%C_{16})_{exp}$	$(\%C_{16})_{pred}$
H	40 ± 2	32
o-CH_3	31 ± 3	36
m-CH_3	29 ± 3	36
p-CH_3	35 ± 3	37
o-Cl	44 ± 4	39
m-Cl	40 ± 4	37
p-Cl	37 ± 3	37
o-COOH	22 ± 2	24
p-OCH_2CH_3	41 ± 2	38
p-COOH	22 ± 2	26
o-OCH_3	28 ± 2	29
m-OCH_3	33 ± 5	30
p-OCH_3	41 ± 3	34
Test set	(C_{16}) / %	(T_{16}) / %
2,4-Dichloro	27 ± 6	39
3,4-Fimethoxy	31 ± 3	43
3,4-Fichloro	16 ± 5	30

Hypotriglyceridemic activity

Table 3 shows that the dipole moment (μ) is the best descriptor (R = –0.64) for the structure-hypotriglyceridemic activity relationship in 2-(substituted phenyl)indan-1,3-diones. The second best descriptor is the LUMO energy (R = 0.58). Unfortunately, these descriptors cannot be used together in a classical multiple regression equation since they are highly correlated with R = –0.95. The best regression equation to describe pT_{16} is thus:

$$pT_{16} = 2.10(\pm 0.20) - 0.20(\pm 0.07)\ \mu \quad (5)$$

with n = 13; R = – 0.64, $R^2 = 0.41$, F = 7.74, s = 0.23 and $p = 0.018$. The values in parentheses are the standard error estimates for the regression coefficients. Though the R^2 value is not very impressive, both coefficients are statistically significant at a 95% confidence level. According to this regression equation, 2-(substituted phenyl)indan-1,3-diones possessing lower dipole moments should exhibit better hypotriglyceridemic activity. It was also evaluated how this descriptor is influenced by the conformational effect as it was done for pC_{16}, although in this case there were no statistically significant results.

The use of AM1 descriptors considering the solvent effect simulated by the COSMO method only slightly altered the correlation matrix with respect to the matrix obtained for isolated molecules, as shown in Table 3. As a consequence, the regression equation is similar to equation 5, which was obtained for the isolated molecules using the most stable conformation for each compound. As can be seen below, the correlation coefficient is slightly increased to 0.66:

$$pT_{16} = 2.31(\pm 0.26) - 0.19(\pm 0.06)\ \mu \quad (6)$$

with n = 13; R = – 0.66, $R^2 = 0.44$; F = 8.66, s = 0.23 and $p = 0.013$. The R^2 value is still not very impressive, but both coefficients are statistically significant at a 95% confidence level. Table 6 shows the percentage reduction in hypotriglyceridemic activity levels obtained from equation 6 and the experimental values for the testing set.

Table 6. Reduction percents of hypotriglyceridemic activity levels obtained from equation 6 and from experimental results

Test set	$(\%T_{16})_{exp}$	$(\%T_{16})_{pred}$
2,4-Dichloro	13 ± 5	20
3,4-Dimethoxy	11 ± 4	15
3,4-Dichloro	34 ± 8	17

In general, there is a good fit between them. On the other hand, it is also possible to notice that the agreement for the 3,4-dichloro derivative is not as good.

In addition to the dipole moment and LUMO energy, it was used three other descriptors (qC_{17}, log P and HOMO energy) to perform a PCA for pT_{16}. However, there was no significant improvement in the results compared to those obtained from equations 5 and 6.

A new hypolipidemic agent

Our results reveal that more active compounds in reducing cholesterol levels are expected to be associated with high log P and π values and with low qH_{14} and qC_{17} values, especially the hydrogen atomic charge bonded to the sp^3 carbon of the indan group, as shown by equations 1 to 3. Our results also reveal that molecules having lower dipole moments should be better at reducing serum triglyceride levels. It was found that the hydroxyl group in the *ortho* position on the phenyl group possesses such characteristics. It was thus performed AM1 calculations on 2-(*o*-hydroxy-phenyl)indan-1,3-dione in order to obtain its qH_{14}, qC_{17} and μ values. The calculated values are 0.143 |e|, –0.085 e^- and 1.828 D, respectively, for the isolated molecule. It was also obtained the values of log P (3.01) and π (–0.67) for the molecule.

Table 7 shows the pC_{16} values for 2-(*o*-hydroxy-phenyl)indan-1,3-dione using equations 2 and 4, as well as its pT_{16} value using equation 5. The table also lists the experimental and calculated pC_{16} and pT_{16} values for the parent compound, 2-phenylindan-1,3-dione (**I**). Compound **I** reduces cholesterol and triglyceride levels by 35 and 40%, while the corresponding reduction values for 2-(*o*-hydroxy-phenyl)indan-1,3-dione are 57 and 54%. Hence, it is concluded that 2-(*o*-hydroxy-phenyl)indan-1,3-dione should be considered for new experimental screening studies on compounds to reduce serum cholesterol and triglyceride levels.

Table 7. Reduction percents of hypolipidemic activity levels for both 2-(*o*-hydroxy-phenyl)indan-1,3-dione and 2-phenylindan-1,3-dione using equations 2 and 4 for hypocholesteremic activities and equations 5 and 6 for hypotriglyceridemic activities. Experimental values of 2-phenyl-indan-1,3-dione are also given in brackets[a]

Compound	$(\%C_{16})$	$(\%T_{16})$
2-(*o*-Hydroxy-phenyl)indan-1,3-dione	57	54
	(32)	(62)
2-Phenylindan-1,3-dione	35	40
	(32)	(46)
	[40 ± 2]	[40 ± 4]

[a]Values obtained from equations 4 and 6 for pC_{16} and pT_{16}, respectively, are given in parentheses.

Conclusions

AM1 electronic descriptors allowed us to build molecular models in order to obtain structure-activity relationships in a group of sixteen 2-(substituted phenyl)indan-1,3-diones possessing hypolipidemic activities. Two different linear models were built, one to describe the hypocholesteremic activity and the other to describe the hypotriglyceridemic activity, since there is not a linear correlation between them. Thus, the hypocholesteremic activity of 2-(substituted phenyl)indan-1,3-diones showed a reasonable linear correlation with the hydrogen atomic charge bonded to the carbon atom of the indan ring, whereas the dipole moment was the most significant descriptor for explaining hypotriglyceridemic activity changes. The statistical analysis supporting such relationships was performed, considering the molecular environment with and without solvent effect. Although the R^2 values were not very impressive, the linear coefficients were statistically significant at the 95% confidence level. Furthermore, the residual analysis also supported this behavior.

From these models, it was proposed new compounds that may be more active than any of the molecules employed in the training and test sets. As a consequence, such compounds may be viewed as potential candidates for new experimental screening studies. From a theoretical point of view, it would be now interesting to employ molecular orbital calculations using a density functional such as B3LYP with an extended basis set including both polarization and diffuse functions in order to test our AM1 molecular modeling.

Acknowledgments

This work is dedicated to Professor Benício de Barros Neto, who introduced the chemometric techniques in the Departamento de Química Fundamental of the Universidade Federal de Pernambuco. We are also grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for a fellowship granted to A. G. S.

References

1. Cechinel Filho, V.; Campos, F.; Corrêa, R.; Yunes, R. A.; Nunes, R. J.; *Quim. Nova* **2003**, *26*, 230.
2. Reynolds, D. J.; Wong, O. T.; Simlot, R.; Chang, J. J.; Hall, I. H.; *Arch. Pharm. (Weinheim, Ger.)* **1994**, *327*, 237.
3. El-Zahabi, M. A.; Gad, L. M.; Bamanie, F. H.; Al-Marzooki, Z.; *Med. Chem. Res.* **2012**, *21*, 75.
4. Chapman, J. M.; Cocolas, G. H.; Hall, I. H.; *J. Med. Chem.* **1979**, *22*, 1399; Wyrick, S. D.; Voorstad, P. J.; Cocolas, G. H.; Hall, I. H.; *J. Med. Chem*. **1984**, *27*, 768; Hall, I. H.; Williams, W. L.; Wyrick, S. D.; Voorstad, P. J.; *J. Pharm. Sci*. **1984**, *73*, 1482; Chapman, J. M.; Voorstad, P. J.; Cocolas, G. H.; Hall, I. H.; *J. Med. Chem*. **1983**, *26*, 237.
5. Dewar, M. J. S.; Zoebisch, E. G; Eamonn, F. H.; Stewart, J. P.; *J. Am. Chem. Soc.* **1985**, *107*, 3902.
6. Ramos, M. N.; Neto, B. B.; *J. Comp. Chem.* **1990**, *11*, 569.
7. Coutinho-Neto, M. D.; Pinheiro Filho, W. C.; Neto, B. B.; Ramos, M. N.; *J. Braz. Chem. Soc.* **1993**, *4*, 139.
8. Murthy, A. R.; Wyrick, S. D.; Hall, I. H.; *J. Med. Chem*. **1985**, *28*, 1591.
9. Koelsch, C. C. F.; Byers, D. J.; *J. Org. Chem.* **1940**, *62*, 560; Koelsch, C. F.; *J. Am. Chem. Soc.* **1936**, *58*, 1331.
10. Mosher, W. A.; Soeder, R. W.; *J. Org. Chem.* **1971**, *36*, 1561.
11. Shapiro, S. L.; Geiger, K.; Freedman, L.; *J. Org. Chem.* **1960**, *25*, 1860.
12. Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Pople, J.; *Gaussian 98W* (*revision A11.2*); Gaussian, Inc.: Pittsburgh, PA, 2001.
13. Kowwin™, v. 1.67; US Environmental Protection Agency, 2000.
14. Jolliffe, I. T.; *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002.
15. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T.; *A User-Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, UK, 2002.
16. Klamt, A.; Schüürmann, G.; *J. Chem. Soc*.**1993**, *2*, 799.
17. Barone, V.; Cossi, M.; Tomasi, J.; *J. Chem. Phys*. **1997**, *107*, 3210.
18. Becke, A. D.; *J. Chem. Phys*. **1993**, *98*, 5648.

Submitted: June 19, 2012
Published online: September 21, 2012