Journal of Enzyme Inhibition and Medicinal Chemistry, 2011; 26(1): 78–87
Copyright © 2011 Informa UK, Ltd.
ISSN 1475-6366 print/ISSN 1475-6374 online
DOI: 10.3109/14756361003724760

informa
healthcare

RESEARCH ARTICLE

Prediction of the relationship between the structural features of andrographolide derivatives and α-glucosidase inhibitory activity: A quantitative structure-activity relationship (QSAR) Study

NS Hari Narayana Moorthy, Maria J Ramos, and Pedro A Fernandes

REQUIMTE, Department of Chemistry, Faculty of Sciences, University of Porto, Street of Campo Alegre, Porto, Portugal

Abstract
In order to predict the structural features responsible for α-glucosidase inhibitory activity, a quantitative structure-activity relationship (QSAR) analysis was performed on a series of andrographolide derivatives. To determine the quantitative relationship for the statistically significant models in terms of r (>0.8), F (99%) and Q^2 (>0.71) values were selected. The promising results we obtained could be used to predict the structural requirements for the inhibition of α-glucosidase activity. The models developed included: subdivided surface area, adjacency, surface volume and shape, molecular orbital package (MOPAC) and partial charge descriptors and showed a high correlation with the inhibitory activity. The descriptors used revealed that a van der Waals (vdW) surface with significant polar volume is favourable to the activity. The positive effect of the shape descriptors; PM3-LUMO and vsurf_wp7 and the negative effect of GCUT_PEOE_2 indicated that the active site may contain some nucleophilic positions that could interact with the ligand and the hydrogen acceptor and/or donor groups for hydrogen bonding with inhibitors.

Keywords: α-glucosidase inhibitory activity, andrographolide, volsurf, QSAR, petitjean

Introduction

The enzyme α-glucosidase (EC 3.2.1.20) belongs to the glucohydrolyse group of enzymes, and catalyses the final step in the digestive process of carbohydrates which is essential for glycoprotein biosynthesis. The members of this group share the ability to release the terminal glucose moiety from the non-reducing end of their substrate. The inhibitors of α-glucosidase are known to possess a large number of therapeutic effects, including antitumour, anti-diabetes, antiviral and immunoregulatory agent. [1–9].

The human immunodeficiency virus (HIV) envelope contains two glycoproteins, gp120 and gp41, these bind to the cell surface receptor CD4 for fusion of the viral and cellular membranes by way of a conserved fusion domain in the N-terminal region. Nowadays, these targets play an important role in the development of new anti-HIV drugs, due to the fact that inhibitors of this enzyme retard the carbohydrate digestion and/or inhibit the biosynthesis of the N-linked oligosaccharides on the envelope glycoprotein [10–15].

Despite the importance of α-glucosidase inhibition, it has not been possible to obtain the experimental structure of α-glucosidase and this has limited the understanding of the nature of its interaction with inhibitors. The homology modelled enzyme has provided some useful information for the creation of new biologically active molecules [16,17]. In the present investigation, we describe the structural features required for α-glucosidase inhibitory activity by quantitative structure-activity relationship (QSAR) analysis.

The aim of a QSAR analysis is to investigate the correlation between activity (usually biological activity)

Address for Correspondence: Pedro A Fernandes, REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, Porto, Portugal. Tel: +351-220 402 667; Fax: +351-220 402 659; E-mail: pafernan@fc.up.pt

(Received 06 October 2009; revised 12 February 2010; accepted 23 February 2010)

and the properties of a set of molecules [18]. With this strategy in mind, a series of andrographolide derivatives with α-glucosidase inhibitory activity were considered in order to find the relationship between their structural properties and biological activity [9]. Earlier studies in our laboratory on a structurally different series of α-glucosidase inhibitors showed that the electrotopological, molecular connectivity and hydrophobicity parameters are responsible for the α-glucosidase inhibitory activity. In the present investigation, we utilised some additional parameters such as partial charge, electronic, subdivided surface area and polar surface volume, and shape descriptors, to correlate the α-glucosidase inhibitory activity with the structure of the andrographolide derivatives.

Experimental

Dataset

The molecular structure and the α-glucosidase inhibitory activity values of 25 andrographolide derivatives were obtained from the published work of Xu et al. [19] Only 19 compounds in the series have defined activities and the remaining 6 compounds do not inhibit at a concentration of 100 μM (Table 1). The former 19 compounds were considered for the study by converting the half maximal inhibitory activity (IC_{50}) value into $\log 1/IC_{50}$ or $-\log IC_{50}$ (pIC_{50}), which is proportional to the free energy changes. Some compounds in the series have inhibitory activities reported as a percentage inhibition converted as pIC_{50} by log (P/100-P), where P is the percentage inhibition [20].

Computational study: descriptor calculation

In the present investigation, Chemdraw ultra module of ChemBioOffice 2008 software (Cambridgesoft, Cambridge, UK) [21] was used to draw the structures of the dataset. The semi empirical molecular orbital package MOPAC program with Hamiltonian Austin Model 1 (AM1) force field with 0.05 RMS gradients of Molecular Operating Environment (MOE) software (Chemical Computing Group, Montreal, Canada) [22] was used to optimise the lower energy geometry of the molecules. This package has been extended to calculate the 2D and 3D descriptors of the molecules. A large number of theoretical molecular descriptors are available in the package to define the structural properties of molecules explicitly.

Computational study: statistical analysis

In order to quantify the correlation, QSAR models were developed using observed α-glucosidase inhibitory activities as the dependent variables and the calculated physicochemical descriptors as the independent variables for multiple linear regression analysis. Statistica 8.0 software (StatSoft, Tulsa, OK, USA) [23] was used to develop statistically significant models for the complete data set that possessed defined activity. In order to reduce the redundant and useless information, descriptors that possess zero correlation to the dependent variable (biological activity) as well as descriptors showing intercorrelation superior to 0.6 were discarded (the correlation matrix is provided in Table 2). This rule was adopted for a limiting number of descriptors in the model [24]. The significant models were selected for further study taking into account the high correlation coefficients, F_{test}, t_{test} values and the significance of the descriptors included in the model building.

Multicollinearity is an important factor that influences the QSAR models. To confirm the absence of multicollinearity, the variable inflation factor (VIF) was calculated for each descriptor in the regression. VIF denotes the fact that the variance of the standardised regression coefficients can be computed as the product of the residual variance (for the correlation transformed model) [18,25] as shown in Table 3.

The Durbin-Watson (DW) test was employed to check the serial correlation residuals (correlation of adjacent residuals), i.e. whether residuals for adjacent cases are correlated, indicating that the observations or cases in the data file are not independent [26] as shown in Table 3.

In order to determine the stability and reliability of the selected models, a validation was performed by the leave one out (LOO) method. A high Q2 (for instance $Q2 > 0.5$) may be considered as an indicator, or even as the ultimate proof, that the model is predictive [27–30]. The significant cross validated correlation coefficient and Cook`s distances of the compounds provide evidence for the stability and reliability of the models (Table 4).

Cook's distances indicate the distance between the computed B (regression coefficient) values and the values one would have obtained after the respective case had been excluded. All distances should be of about equal magnitude, otherwise there is reason to believe that the respective case may have biased the estimation of the regression coefficients [31–33].

Results and discussion

Quantitative structure activity relationship analysis of a series of andrographolide derivatives were performed. We have initially explored over 20 models (provided as supplementary materials). However, most of them suffered from the statistical parameters (r, F_{test}, t_{test}, Q^2, multicollinearity, autocorrelation, etc.) and only the statistically significant models are presented and discussed here. In the following models, N is the number of compounds, R is the correlation coefficient, R^2 is the squared correlation coefficient, F stands for Fischer test, t for student t_{test} and Q^2 is the crossvalidated correlation coefficient. The numbers within the parentheses following the coefficient terms are the standard errors of the regression terms. The numbers within the parentheses following the F_{test} and the t_{test} are the tabulated values at the mentioned significance level. Beta value stands for the contribution of each descriptor in the models for the activity prediction.

Table 1. Structure and α-glucosidase inhibitory activity of the andrographolide derivatives.

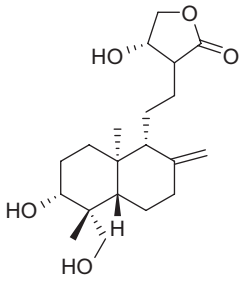
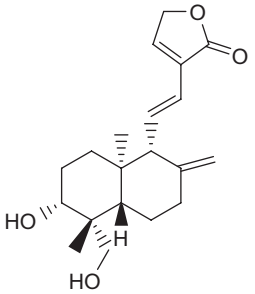
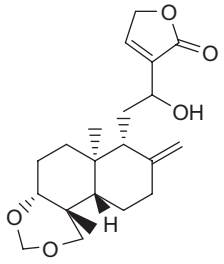
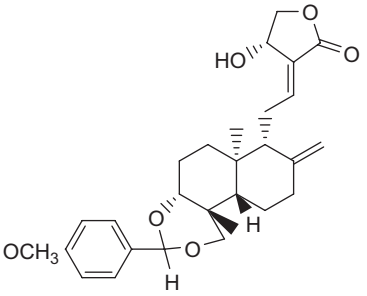
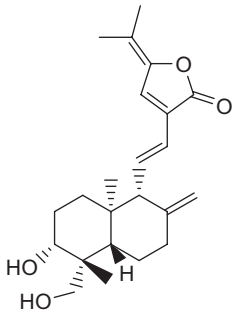
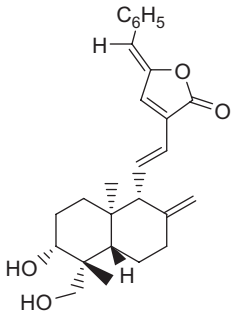
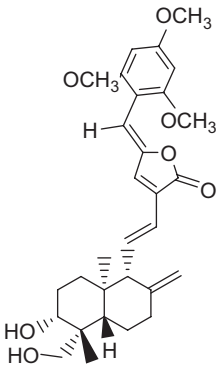
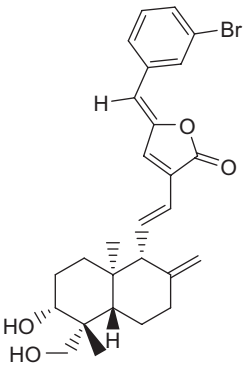
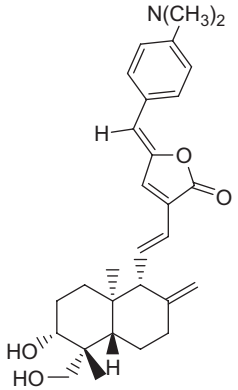
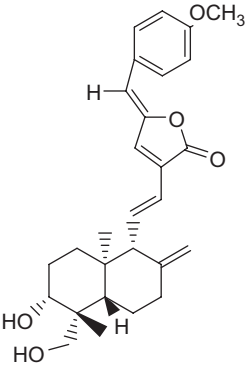
Comp code	Structure	% Inhibition /IC_{50} μM	Comp Code	Structure	% Inhibition /IC_{50} μM
AT1		34	AT2		16.5
AT3		6.9	AT4		13.2
AT5		17.1	AT6		58
AT7		84	AT8		16.7
AT9		70	AT10		16

Table 1. continued on next page

Table 1. Continued.

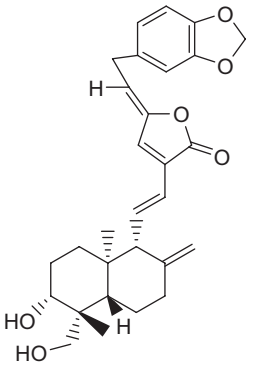
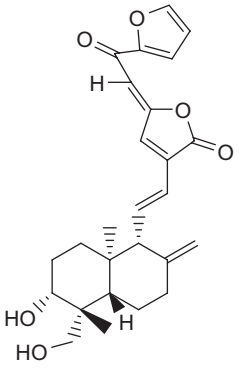
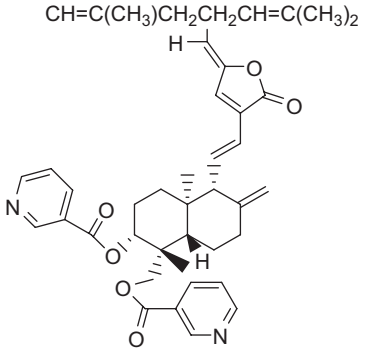
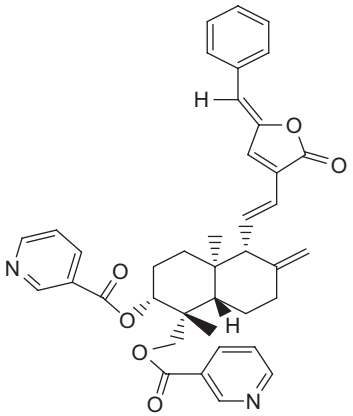
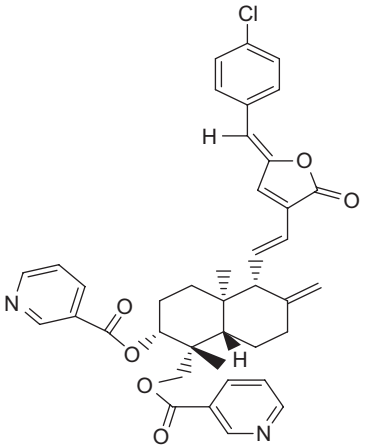
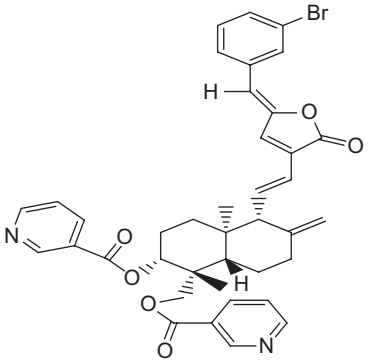
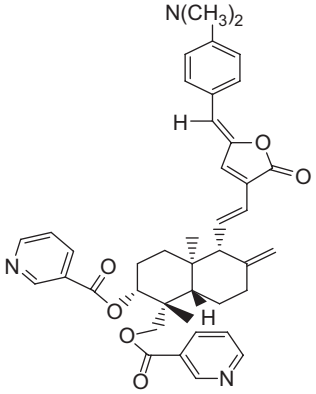
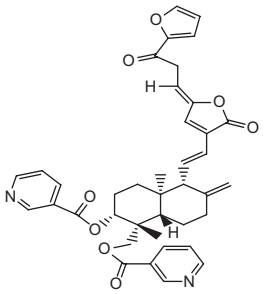
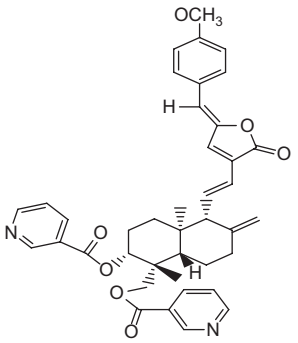
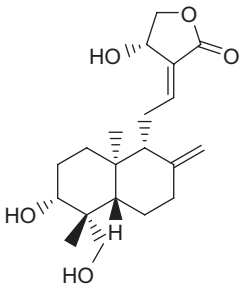
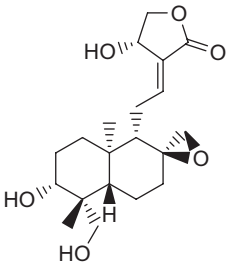
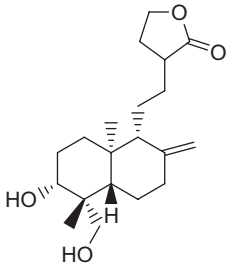
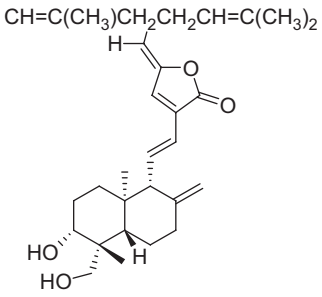
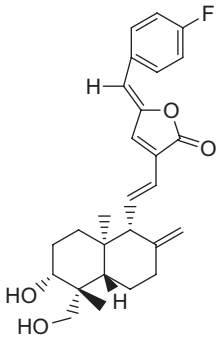
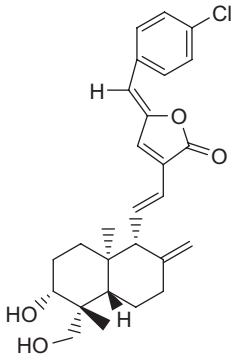
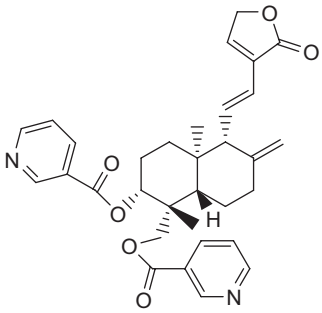
Comp code	Structure	% Inhibition /IC_{50} μM	Comp Code	Structure	% Inhibition /IC_{50} μM
AT11		82	AT12		100
AT13		28	AT14		16
AT15		6	AT16		14
AT17		25	AT18		36

Table 1. continued on next page

Table 1. Continued.

Comp code	Structure	% Inhibition /IC_{50} μM	Comp Code	Structure	% Inhibition /IC_{50} μM
AT19		11	A		NIa
B		NI	C		NI
D		NI	E		NI
F		NI	G		NI

Note: % Inhibition determined at 100 μM concentration.

Model 1

$pIC_{50} = 0.0236\ (\pm\ 0.0061)$ SlogP_VSA2 + 5.3135 (± 0.8683) vsurf_wp7 – 2.6955 (± 0.3431)

Where N = 19, R = 0.8455, $R^2 = 0.7149$, $AdjR^2 = 0.6792$, $Q^2 = 0.7148$, $F_{(2,\ 16,\ 0.01)} = 20.0557$ (6.226), SEE = 0.285, $t_{(0.0005,16)} = -7.856$ (4.015), p = 0.0000, Beta value for SlogP_VSA2 = 0.56 and vsurf_wp7 = 0.886

Model 1 was built with two parameters; SlogP_VSA2, which is a subdivided surface area descriptor, and vsurf_wp7, which is a polar surface volume and shape parameter. The subdivided surface area descriptor describes an approximate accessible van der Waals surface area (in Å^2) calculation for each atom, v_i, along with another atomic property, p_i. The v_i were calculated using a connection table approximation. SlogP_VSA2 is defined to be the sum of the v_i over all atoms *i*. p_i denotes the contribution to logP (o/w) for atom *i* as calculated in the SlogP descriptor, calculated in a specified range, from –0.2 to 0. The partition coefficient (logP) of small molecules can be calculated as the sum of the contribution of each of the atoms in the molecule,

$$P_{calc} = \sum n_i a_i \quad (1)$$

where P_{calc} is the property to be calculated (logP), n_i is the number of atoms i present in the molecule and a_i is the contribution for atoms i [22,34]. The SlogP_VSA2 descriptor contributes positively to the α-glucosidase inhibitory activity. It reveals if the presence of hydrophobicity of the molecules is favourable for the α-glucosidase inhibitory activity.

The Vsurf_wp7 is a 3D descriptor describing the polar volume on the surface area. The vsurf descriptors are similar to the VolSurf descriptors and are useful for pharmacokinetic property prediction. The vsurf_wp7 descriptor depends on the structure, connectivity and conformation (with the dimensions measured in Å) at the -5.0 kcal/mol energy level and may be defined as the molecular envelope accessible to the water molecules. The volume of this envelope varies with the level of interaction energies between the water and the

Table 2. Correlation matrix for the descriptors contributed in the selected QSAR models.

Descriptor	Petitjean	GCUT_PEOE2	Q_VSA_FPNEG	PM3_LUMO	SlogP_VSA2	Vsurf_Wp7	pIC_{50}
Petitjean	1.0000	0.3897	0.0731	0.1491	0.1433	0.4474	0.4929
GCUT_PEOE2	0.3897	1.0000	0.0752	−0.4073	−0.1929	−0.0938	−0.3377
Q_VSA_FPNEG	0.07317	0.0752	1.0000	0.2231	−0.2112	0.5149	0.0902
PM3_LUMO	0.1491	−0.4073	0.2231	1.0000	0.0668	0.6623	0.6339
SlogP_VSA2	0.1433	−0.1929	−0.2112	0.0668	1.0000	−0.3866	0.2176
Vsurf_Wp7	0.4474	−0.0938	0.5149	0.6623	−0.3866	1.0000	0.6693
pIC_{50}	0.4929	−0.3377	0.0902	0.6339	0.2178	0.6693	1.0000

Table 3. Redundancy (Tolerance and VIF) and Durbin-Watson value of the models and descriptors.

Model	Variable	Tolerance	R^2	VIF	D-W	D-W (Tab)
Model 1	SlogP_VSA2	0.8506	0.1494	1.1757	2.0073	(0.967–1.685)
	Vsurf_Wp7	0.8506	0.1494	1.1757		
Model 2	Q_VSA_FPNEG	0.7348	0.2652	1.3610	1.8557	(0.859–1.848)
	SlogP_VSA2	0.8504	0.1496	1.1760		
	Vsurf_Wp7	0.6541	0.3459	1.5287		
Model 3	Petitjean	0.9782	0.0218	1.0223	1.6899	(0.897–1.710)
	PM3_LUMO	0.9782	0.0218	1.0223		
Model 4	Petitjean	0.6997	0.3003	1.4292	1.7586	(0.779–1.900)
	GCUT_PEOE2	0.6068	0.3932	1.6479		
	PM3_LUMO	0.7455	0.2545	1.3414		

DW (Tab): Tabulated Durbin-Watson values at 5% significance level.

Table 4. Predicted activity and Cook's distance for the molecules.

		Model 1		Model 2		Model 3		Model 4	
Comp code	OA	PA	CD	PA	CD	PA	CD	PA	CD
AT1	−0.29	−0.24	0.013	−0.10	0.2291	0.02	0.4881	−0.27	0.0936
AT 2	−0.7	−0.39	0.4546	−0.64	0.0449	−0.86	0.0227	−0.93	0.0779
AT 3	−1.13	−1.05	0.0068	−1.22	0.0157	−1.45	0.3282	−1.08	0.019
AT 4	−0.82	−1.05	0.0514	−0.97	0.0269	−1.09	0.1094	−0.75	0.0188
AT 5	−0.69	−1.06	0.0974	−1.06	0.0968	−1.06	0.1298	−0.66	0.0024
AT 6	−1.76	−1.72	0.0017	−1.71	0.0025	−1.68	0.0052	−1.94	0.0786
AT 7	−0.7	−1.06	0.091	−0.81	0.0305	OL		−0.99	0.3426
AT 8	−1.92	−1.72	0.0358	−1.64	0.0799	−1.75	0.027	−1.96	0.0026
AT 9	−1.85	−1.72	0.0135	−1.57	0.1025	−1.58	0.0381	−1.58	0.0643
AT 10	−1.2	−1.72	0.226	−1.67	0.1946	−1.6	0.0901	OL	
AT 11	−1.91	−1.72	0.0322	−1.88	0.0014	−1.55	0.0702	−1.67	0.0665
AT 12	−2	−1.72	0.0671	−1.96	0.0046	−1.83	0.0292	−2.06	0.011
AT 13	−1.45	−1.2	0.0299	−1.21	0.0273	−1.43	0.0001	−1.37	0.0038
AT 14	−1.2	−1.2	0	−1.17	0.0006	−1.17	0.001	−1.15	0.0044
AT 15	−0.78	−1.2	0.0869	−1.14	0.0713	OL		OL	
AT 16	−1.15	−1.2	0.0014	−1.12	0.0003	−1.23	0.0099	−1.19	0.0035
AT 17	−1.39	−1.2	0.0193	−1.07	0.082	−1.14	0.0716	−1.13	0.1234
AT 18	−1.56	−1.2	0.0623	−1.44	0.0239	−1.51	0.0011	−1.7	0.0245
AT 19	−1.04	−1.2	0.0123	−1.14	0.0056	−1.15	0.014	−1.14	0.0171

OA, observed activity; PA, predicted activity; CD, Cook's distance; OL, outlier.

solute molecule. The Wp7 accounts for the polar and the hydrogen bond donor-acceptor regions (hydrophilic region) [35–37]. The hydrophilic regions are defined as the molecular envelope that is accessible to attract water molecules. The positive contribution of this descriptor shows that an increase in the hydrogen donor and acceptor (hydrophilic) properties on the surface of the molecule increases the α-glucosidase inhibitory activity, and it also reveals that some hydrogen bond donor or acceptor groups may be present on the surface of the enzyme.

Model 2

pIC_{50} = 0.0234 (± 0.0053) SlogP_VSA2 + 6.3429 (± 0.8601) vsurf_wp7 – 7.0389 (± 2.8261) Q_VSA_FPNEG –1.5698 (± 0.5414)

N = 19, R = 0.8935, R^2 = 0.7983, AdjR^2 = 0.7579, Q^2 = 0.7983, $F_{(3,15,0.01)}$ = 19.7864 (5.417), SEE = 0.24755, $t_{(0.01,15)}$ = -2.899 (2.6025), p = 0.011, Beta value for Q_VSA_FPNEG = -0.34, SlogP_VSA2 = 0.555 and vsurf_wp7 = 1.06.

Model 2 is a triparametric equation, with an additional negative signed Q_VSA_FPNEG descriptor along with the subdivided surface area and polar surface, volume and the shape descriptors discussed in model 1. The Q_VSA_FPNEG is a partial charge descriptor that defines the fractional negative polar van der Waals surface area. It calculates the sum of v_i (van der Waals surface area of atom i) in which q_i (partial charge of atom i) is less than -0.2 divided by the total surface area. The v_i were calculated using a connection table approximation in the same way for SlogP_VSA2. The descriptors prefixed with Q_ use the partial charges stored with each structure in the database [22,38]. The negative contribution of the descriptor reveals that the fractional negative charge on the van der Waals surface of the molecules is detrimental to the α-glucosidase inhibitory activity.

Model 3

pIC_{50} = 15.4736 (± 4.0275) Petitjean + 0.5665 (± 0.1172) PM3-LUMO – 8.1253 (± 1.9651)

N = 17, R = 0.8723, R^2 = 0.761, AdjR^2 = 0.7268, Q^2 = 0.761, $F_{(2,14,0.01)}$ = 22.2851, (6.515), SEE = 0.2611, $t_{(0.005,14)}$ = –4.135 (2.9768), p = 0.001, Beta value for Petitjean = 0.508 and PM3-LUMO = 0.638.

This model was constructed with positively contributed descriptors, namely an adjacency and distance matrix descriptor; Petitjean, and a MOPAC descriptor; PM3-HOMO. According to the Petitjean definition, the eccentricity of a vertex corresponds to the distance from that vertex to the most remote vertex in the graph. The distance is obtained from the distance matrix as the count of edges between the two vertices. If r_i is the largest matrix entry in row i of the distance matrix D, then the radius is defined as the smallest of the r_i. The graph diameter D is defined as the largest vertex eccentricity in the graph. The Petitjean number is the value of the (diameter - radius)/diameter and it describes the geometrical shape (flexibility) of the molecules [39,40]. The positive contribution reveals that if the compounds have optimum geometry (flexibility) they will have a favourable α-glucosidase inhibitory activity.

The PM3_LUMO is the energy (eV) of the lowest unoccupied molecular orbital (LUMO) calculated using the PM3 Hamiltonian of the MOPAC package. It is a popular quantum mechanical descriptor, which plays a major role in governing many chemical reactions and the energy of the LUMO is directly related to the electron affinity, characterising the susceptibility of the molecule towards attack by nucleophiles (the unoccupied MO that can accept endogenous electrons). It is a global molecular property that describes the electrophilicity of a compound in general terms, and is a measurement of the ability of the molecule to act as an electron acceptor with lower LUMO values leading to stronger electrophilicity [41–43]. The positive contribution of this descriptor explains that an increase in the LUMO value leads to a decrease in the strength of the electrophilicity of the molecule, which is favourable for the activity. The above mentioned discussion reveals that the activity site of the enzyme may contain low affinity nucleophiles.

Model 4

pIC_{50} = 23.7347 (± 3.2742) Petitjean + 0.3588 (± 0.0928) PM3-LUMO – 5.501 (± 1.0460) GCUT_PEOE_2 – 12.0917 (± 1.5941)

N = 17, R = 0.9504, R^2 = 0.9032, AdjR^2 = 0.8809, Q^2 = 0.9032, $F_{(3,13,0.01)}$ = 40.4463 (5.739), SEE = 0.1795, $t_{(0.0001,13)}$ = –7.585 (4.2208), p = 0.0000, Beta value for Petitjean = 0.748, GCUT_PEOE_2 = –0.58 and PM3-LUMO = 0.386.

This model has an additional GCUT descriptor, GCUT_PEOE_2 contributes negatively to the α-glucosidase inhibitory activity along with a favourable adjacency and distance matrix descriptor (Petitjean) and the MOPAC descriptors (PM3-LUMO). GCUT descriptors uses graph distances instead of bond order information (like BCUTs) for the calculations. They are calculated from the eigenvalues of the modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(d_{ij}) where d_{ij} is the (modified) graph distance between atoms i and j. The diagonal takes the value of the partial equalisation of orbital electronegativities (PEOE) partial charges [22]. The PEOE) is a method of calculating atomic partial charges, in which charge is transferred between bonded atoms until equilibrium. The amount of charge transferred at each iteration is damped with an exponentially decreasing scale factor to guarantee convergence. The amount of charge transferred, dq_{ij}, between atoms i and j when $X_i > X_j$ is shown in Equation 2:

$$dq_{ij} = (1/2k)(X_i - X_j)/X_j^+ \quad (2)$$

Where X_j^+ is the electronegativity of the positive ion of atom j, X_i is the electronegativity of atom i (quadratically dependent on partial charge) and k is the iteration number of the algorithm. PEOE is the electronegativity concept as per equation 3,

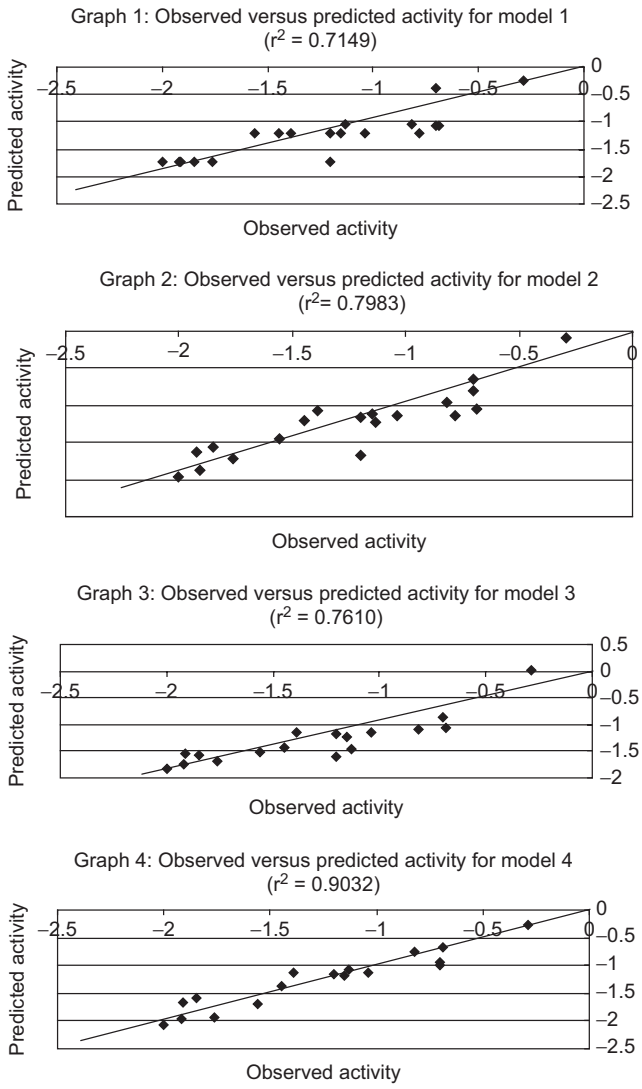


Figure 1. Scatter plot between observed and predicted for models 1–4.

$$\chi_v = 1/2(I_v + E_v) \quad (3)$$

In this equation, the electronegativity is related to its ionisation potential I, and its electron affinity E. The electronegativity of an atom further depends on the charge of other atoms in this orbital and also the charge of the same atom in other orbitals [22,44,45]. The negative contribution of this descriptor shows that the partial negative charge of the molecules is detrimental to the α-glucosidase inhibitory activity. It suggests that the active site of the enzyme may have some electronegative groups for interaction, which is evidenced by the MOPAC descriptors suggesting that the active site of the enzyme may have some nucleophilic groups for interaction.

Statistical analysis of the models

The selected significant QSAR models are biparametric and triparametric equations, mainly having subdivided surface area, electronic and polar surface volume and shape descriptors. In this study, the 3D descriptors along with the 2D descriptors were considered to investigate the role of the structural features of the ligands that correlated with the α-glucosidase inhibitory activity. Models 1 and 2 have been developed for 19 compounds, while we have only considered 17 compounds for models 3 and 4. In the two latter models, AT-15 along with AT-7 in model 3, and AT-10 in model 4, were outliers, because the data points showed residual values of a large magnitude (a value is considered as an outlier when the residual values exceeds twice the standard error of estimate of the model) [46].

The correlation coefficient (r) explains the variation in the observed data (experimental) and its value varies from –1 to +1. The closer the R value is to 1, the better the fit of the regression equation. The significant models had correlation coefficient values larger than 0.84 and provided significant variation in biological activity in terms of the squared correlation coefficient. This confirms that the selected descriptors in the models greatly contributed to the activity prediction. The F_{test} value of the model was larger than the tabulated value at 99% significance. The t_{test} values of the models have significance at 0.0001 and 0.005 confidence levels and the values were larger than the tabulated values (large margin of difference), which shows that the models are statistically significant and this is important for further study.

The variance inflation factor provides information regarding the multicollinearity of the contributed descriptors in the regression models. A value greater than 10 is an indication of potential multicollinearity problems. A VIF of 10 or even one as low as 4 (equivalent to a tolerance level of 0.10 or 0.25) has been used to indicate excessive or serious multicollinearity. The VIF value of the descriptors in the models were <1.65, which showed that the models are free from multocollinearity [18,24,25]. The VIF values of the models are presented in Table 3.

Durbin-Watson (D-W) statistics are useful for evaluating the presence or absence of a serial correlation of residuals [26]. The value of *d* always lies between 0 and 4. If the DW statistic is substantially less than 2, there is evidence of positive serial correlation and a value close to 4 indicates negative autocorrelation. In the present study, the D-W values of the derived models were greater than 1.68, which show that the values are above the level for positive autocorrelation and below the level for negative autocorrelation of the tabulated upper and lower bound values at a 5% significance level (Table 3).

The statistically significant models derived from the study were validated by the leave-one-out method. The cross validated correlation coefficient (Q^2) value for the models examines their self consistency, which implies a quantitative assessment of the model's robustness and its predictive power. In the present selected models, the Q^2 is >0.71, which showed that the models have a significant predictive power. It may be considered that a $Q^2 > 0.5$ is an indicator that the models have a sufficient predictive power and self consistency [28–30]. In all the models, the residual value between the observed and predicted

activity was considerably low. The correlation between the observed and predicted activity is represented graphically in Figure 1 (Graph 1–4). These outlier compounds were considered for validation by the selected models (models 3 and 4) as test set compounds and the obtained results are given in Table 4. The results show that the models predict the activity of the outlier compounds with the least residual values.

The distance based approaches are also a a method of validation for the models, calculating the distance from each point to a particular point of the data set. The average Cook's distance value of the models is <0.4, which is <1 (squared Cook's distance) [31–32] and the Cook's distance for all the compounds have an almost equal magnitude (<1), which shows that the models have a significant predictive ability for α-glucosidase inhibitory activity. The predicted activity along with the Cook's distance is given in Table 4.

Conclusion

In the present investigation, a series of andrographolide derivatives were considered in order to investigate the quantitative correlation between various descriptors (subdivided surface area, electronic, MOPAC, polar surface volume and shape) with α-glucosidase inhibitory activity. The QSAR results obtained from the regression analysis showed that the selected models were statistically significant. The models had correlation coefficients greater than 0.84, which showed the goodness of fit of the selected significant models. The F_{test} value exceeded the tabulated value at a significance of 99%. The models gave satisfactory Q^2 values and Cook's distance, revealing that the selected models were reliable and possess a high predictive power. The descriptors used in the models were free from multicollinearity and also were free from serial correlation. The descriptors included in the models showed that the subdivided surface area descriptor SlogP_VSA2, polar surface volume and shape descriptors; vsurf_wp7, Petitjean and PM3-LUMO all positively contribute to the activity. The partial charge descriptor and GCUT descriptors negatively contribute to the activity. These descriptors revealed that the compounds with an optimum partition coefficient and polar surface volume in the van der Waals surface are favourable to α-glucosidase inhibitory activity while the LUMO values showed that the active site may have some nucleophilic positions for interaction. This was also confirmed by the PEOE descriptor that is detrimental for the α-glucosidase inhibitory activity. The beta coefficient of the descriptors in the models suggested that the vsurf_wp7 made a significant contribution to the α-glucosidase inhibitory activity prediction. The active site may have some hydrogen bond donor and/or acceptor groups for hydrogen bonding with the inhibitor/water molecules and also the inhibitors should have some flexible structure for an electrophilic interaction. In fact it is possible to propose new molecules from this study. However, in our group this present study will be fundamental for the built up of the homology modelled enzymes for any follow up research on α-glucosidase inhibitors in our laboratory. We will continue to work on protein-ligand docking, molecular dynamic simulations, and detailed free energy calculation studies that may lead to the design of novel α-glucosidase inhibitors for AIDS therapy.

Acknowledgements

One of the authors N.S.H.N. Moorthy is grateful to the Fundaçao para a Ciencia e Tecnologia (FCT), Portugal, for a Postdoctoral Grant (SFRH/BPD/44469/2008).

Declaration of interest

The authors gratefully acknowledge FCT for financial support for project PTDC/QUI/68302/2006. The authors report no declarations of interests. The authors alone are responsible for the content and writing of the paper.

References

1. Pili R, Chang J, Partis RA, Mueller RA, Chrest FJ, Passaniti A. The α-glucosidase I inhibitor castanospermine alters endothelial cell glycosylation, prevents angiogenesis, and inhibits tumor growth. Cancer Res 1995;55:2920–2926.
2. Humphries MJ, Matsumoto K, White SL, Olden K. Inhibition of experimental metastasis by castanospermine in mice: Blockage of two distinct stages of tumor colonization by oligosaccharide processing inhibitors. Cancer Res 1986;46:5215–5222.
3. Papandreou MJ, Barbouche R, Guieu R, Kieny MP, Fenouillet E. The α-glucosidase inhibitor 1-deoxynojirimycin blocks human immunodeficiency virus envelope glycoprotein-mediated membrane fusion at the CXCR4 binding step. Mol Pharmacol 2002;61:186–193.
4. Ouzounov S, Mehta A, Dwek RA, Block TM, Jordan R. The combination of interferon α-2b and n-butyl deoxynojirimycin has a greater than additive antiviral effect upon production of infectious bovine viral diarrhea virus (BVDV) in vitro: Implications for hepatitis C virus (HCV) therapy. Antiviral Res 2002;55:425–435.
5. Kameda Y, Asano N, Yoshikawa M. Valiolamine, a new α-glucosidase inhibiting aminocyclitol produced by *Streptomyces hygroscopicus*. J Antibiotics 1984;37:1301–1307.
6. Robinson KM, Begovic ME, Rhinehart BL, Heineke EW, Ducep JB, Kastner PR, Marshall FN, Danzin C. New potent α-glucohydrolase inhibitor MDL 73945 with long duration of action in rats. Diabetes 1991;40:825–830.
7. Fujisawa T, Ikegami H, Inoue K, Kawabata Y, Ogihara T. Effect of two α-glucosidase inhibitors, voglibose and acarbose, on postprandial hyperglycemia correlates with subjective abdominal symptoms. Metab Clinic Exp 2005;54:387–390.
8. Van Den Broek LAGM, Kat-Van Den Nieuwenhof MWP, Butters TD, Van Boeckel CAA. Synthesis of α-glucosidase I inhibitors showing antiviral (HIV-1) and immunosuppressive activity. J Pharm Pharmacol 1996;48:172–178.
9. Gui FD, Hai WX, Jun FW, Feng WL, Hong ML. Studies on the novel α-glucosidase inhibitory activity and structure–activity relationships for andrographolide analogues. Bioorg Med Chem Lett 2006;16:2710–2713.
10. Silva CHTP, Taft CA. Computer aided molecular design of novel glucosidase inhibitors for AIDS treatment. J Biomol Struct Dyn 2004;22:59–63.
11. Fischer PB, Colin M, Karlson GB, James W, Butters TD, Davis SJ, Gordon S, Dwek RA, Platt FM. The α-glucosidase inhibitor N-butyldeoxynojirimycin inhibits human immunodeficiency virus entry at the level of post-CD4 binding. J Virol 1995;69:5791–5797.
12. Taylor DL, Kang MS, Brennan TM, Bridges CG, Sunkara PS, Tyms AS. Inhibition of α-glucosidase I of the glycoprotein-processing enzymes by 6-O-butanoyl Castanospermine (MDL 28,574) and its

consequences in human immunodeficiency virus infected T cells. Antimicrob Agents Chemother 1994;38:1780–1787.
13. Jung-Hum P, Sungmin K, Hwangseo P. Toward the virtual screening of α-glucosidase inhibitors with the homology-modelled protein structure. Bull Korean Chem Soc 2008;29:921–927.
14. Hwangseo P, Kyo YH, Young HK, Kyung HO, Jae YL, Keun K. Discovery and biological evaluation of novel α-glucosidase inhibitors with *in vivo* antidiabetic effect. Bioorg Med Chem Lett 2008;18:3711–3715
15. Rawling AJ, Lomas H, Adam WP, Marvin JRL, Dominic SA, Shane JSR, Sarah FJ, George WJF, Raymond AD, John HJ, Terry DB. Synthesis and biological characterization of novel N-alkyl deoxynojirimycin α-glucosidase inhibitors. ChemBioChem 2009;10:1101–1105.
16. Kavitha B, Nagakumar B, Ki HP, Keun WL. Binding mode analyses and pharmacophore model development for sulphonamide chalcone derivatives, a new class of α-glucosidase inhibitors. J Mol Graph Model 2008;26:1202–1212.
17. Tomich CH, da Silva P, Ivone C, Taft CA. Homology modelling and molecular interaction field studies of α-glucosidase as a guide to structure based drug design of novel proposed anti-HIV inhibitors. J Comput Aided Drug Des 2005;19:83–92.
18. Karthikeyan C, Moorthy NSHN, Trivedi P. QSAR study of substituted 2-pyridinyl quanidine as selective urokinase-type plasminogen activator (uPA) inhibitors. J Enz Inhib Med Chem 2009;24:6–13.
19. Xu HW, Dai GF, Liu GZ, Wang JF, Liu HM. Synthesis of andrographolide derivatives: A new family of α-glucosidase inhibitors. Bioorg Med Chem 2007;15:4247–4255.
20. Prasanna S, Manivannan E, Chaturvedi SC. Quantitative structure-activity relationship analysis of a series of 2,3-diaryl benzopyran analogues as novel selective cyclooxygenase-2 inhibitors. Bioorg Med Chem Lett 2004;14:4005–4011.
21. Chem BioOffice 2008, CambridgeSoft, Cambridge UK. 2008.
22. MOE is a molecular modelling package developed by Chemical Computing Group, Montreal Canada. 2002.
23. Systatica 8.0 statistical software, StatSoft, Tulsa, OK, USA. 2002.
24. Balaji S, Karthikeyan C, Moorthy NSHN, Trivedi P. QSAR modelling of HIV-1 reverse transcriptase inhibition by benzoxazinones using a combination of P_VSA and pharmacophore feature descriptors. Bioorg Med Chem Lett 2004;14:6089–6094.
25. Cho DH, Lee SK, Kim BT, No KT. Quantitative Structure Activity relationship (QSAR) study of new fluorovinyloxyacetamides. Bull Korean Chem Soc 2001;22:388–394.
26. Durbin J, Watson GS. Testing for serial correlation in least squares regression. II. Biometrika 1951;38:159–178.
27. Tropsha A, Gramatica P, Gombar VK. The importance of earnest, validation is the absolute essential application and interpretation of QSPR models. QSAR Comb Sci 2003;22:69–77.
28. Erikkson L, Jaworsha J, Worth AP, Mark TDC, Robert MM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluation of classification and regression based QSARs. Environ Health Perspect 2003;111:1361–1375.
29. Alexander G, Alexander T. Beware of Q2. J Mol Graphs Model 2002;20:269–276.
30. Gramatica P. Principle of QSAR Model validation: internal and external. QSAR Comb Sci 2007;26:694–701.
31. Cook RD. Influential observation in linear regression. J Am Stat Assoc 1979;74:169–174.
32. Rotriquez G. Chapter 2, Linear models for continuous data, 1993–2000: 49–58 (Lecture Notes http://data.princeton.edu/wws509/notes)
33. Dogra, Shailay K. Script for computing linear regression diagnostics from QSAR world-free online resources for QSAR modelling. (http://www.qsarworld.com/virtual-workshop.php.)
34. Wildman AS, Crippen GM. Prediction of physicochemical parameters by atomic contributions. J Chem Inf Comput Sci 1999;39:868–873.
35. Cruciania G, Crivorib P, Carruptb P-A, Testa B. Molecular fields in quantitative structure–permeation relationships: The VolSurf approach. J Mol Struct (Theochem) 2000;503:17–30.
36. http://www.moldiscovery.com/docs/volsurf/descriptors.html#AEN669
37. Cruciani G, Pastor M, Guba W. Volsurf: a new tool for the pharmacokinetic optimization of lead compounds. Eur J Pharm Sci 2000;11:S29–S39.
38. Shahapurkar S, Kawathekar N, Chaturvedi SC. Quantitative structure activity relationship studies of diaryl thiophen derivatives as selective COX-2 inhibitors. Pharmazie 2005;60:254–258.
39. Petitjean M. Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds. J Chem Inf Comput Sci 1992;32:331–337.
40. Petitjean M, Dubois JE. Topological statistics on a large structural file. J Chem Inf Comput Sci 1990;30:332–343.
41. Iglika L, Andrew PW, Tatiana IN, John CD, Mark TDC. Quantitative structure–activity–activity and quantitative structure–activity investigations of human and rodent toxicity. Chemosphere 2006;65:1878–1887.
42. Balasubramanian N, Avinash D, Vishnukant M. QSAR studies of 4,5-dihydro-4-oxo-3H-imidazo[4,5-c]pyridines as potent angiotensin II receptor antagonists by MLR and NLR analysis. ARKIVOC 2007;i:189–1204.
43. Erol E, Hasan T, Semra G, Selami P, Oral O. A DFT-based QSARs study of acetazolamide/sulfanilamide derivatives with carbonic anhydrase (CA-II) isozyme inhibitory activity. Int J Mol Sci 2007;8:145–155.
44. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 1980;36:3219–3228.
45. Mulliken RS. A new electroaffinity scale; Together with data on valence states and on valence ionization potentials and electron affinities. J Chem Phys 1934;2:782–793.
46. Gnanadesikan R, Kettenring JR. Robust estimate, residuals and outlier detection with multipurpose data. Biometrics 1972;28:81–124.