Journal of Molecular Graphics and Modelling 29 (2010) 363–371

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

3D-QSAR (CoMFA and CoMSIA) and pharmacophore (GALAHAD) studies on the differential inhibition of aldose reductase by flavonoid compounds

Julio Caballero*

Centro de Bioinformática y Simulación Molecular, Facultad de Ingeniería en Bioinformática, Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile

ARTICLE INFO

Article history:
Received 4 June 2010
Received in revised form 20 August 2010
Accepted 21 August 2010
Available online 21 September 2010

Keywords:
Aldose reductase
Flavonoids
QSAR
Pharmacophore
Inhibitory activity
CoMFA
CoMSIA
GALAHAD

ABSTRACT

Inhibitory activities of flavonoid derivatives against aldose reductase (AR) enzyme were modelled by using CoMFA, CoMSIA and GALAHAD methods. CoMFA and CoMSIA methods were used for deriving quantitative structure–activity relationship (QSAR) models. All QSAR models were trained with 55 compounds, after which they were evaluated for predictive ability with additional 14 compounds. The best CoMFA model included both steric and electrostatic fields, meanwhile, the best CoMSIA model included steric, hydrophobic and H-bond acceptor fields. These models had a good predictive quality according to both internal and external validation criteria. On the other hand, GALAHAD was used for deriving a 3D pharmacophore model. Twelve active compounds were used for deriving this model. The obtained model included hydrophobe, hydrogen bond acceptor and hydrogen bond donor features; it was able to identify the active AR inhibitors from the remaining compounds. These *in silico* tools might be useful in the rational design of new AR inhibitors.

1. Introduction

The increase of oxidative stress leads to the progression of diabetes-related complications. Results of physiological and biochemical studies reveal that one of the mechanisms of hyperglycemia-induced oxidative stress is the increased flux of polyol pathway (POP) [1]. The flux of POP under normal levels of glucose is very small; however, when glucose levels are elevated, the flux of POP is enhanced, since there is an overflow of glucose from the normal metabolic pathways such as glycolysis. This results in accumulation of sorbitol, which is unable to cross cell membranes [2]. Accumulation of sorbitol can lead to an osmotic imbalance and may contribute on the progression of diabetic complications, such as cataracts, neuropathy, and nephropathy [3–5].

POP encompasses two enzymes: aldose reductase (AR) and sorbitol dehydrogenase. AR is the first and rate-limiting enzyme of the POP that channels excess glucose to the formation of fructose. Experimental studies have demonstrated the effectiveness of AR inhibitors (ARI) when the flux of POP is enhanced; in this sense, they have been proposed as possible pharmacotherapeutic agents [6]. Several ARIs such as flavonoids, isoflavones, coumarins, stilbenes, rosmarinic acid derivatives, thiazolidinediones, etc., have been reported in the literature [7–11]. Some of the ARIs, such as epalrestat [12], zopolrestat [13], sorbinil [14], minalrestat [15] and zenarestat [16], are in the market, but they have weaker activities and side effects. With these precedents, novel ARIs with improved activity and selectivity are desired.

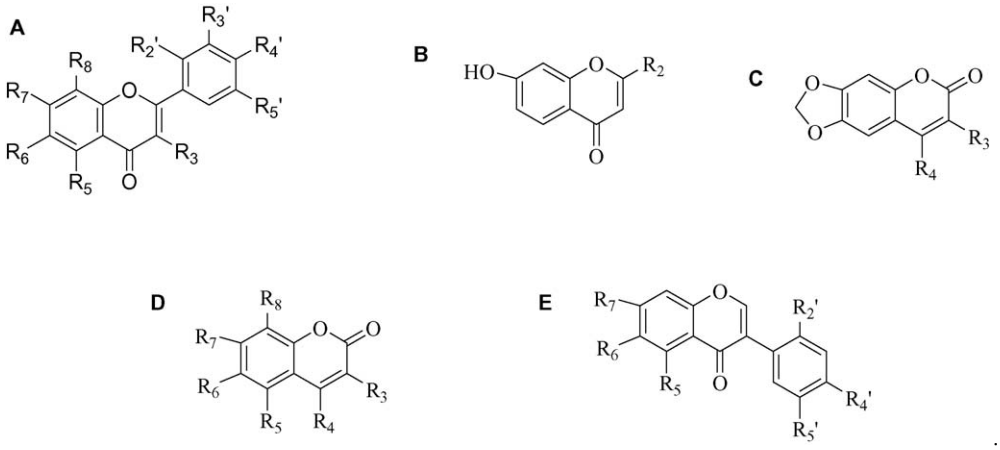
The increment in the speed and efficiency of drug discovery has seen huge investments by major pharmaceutical companies, with the primary aim of reducing cost per synthesized compound or assay. Computational models that are able to predict the biological activity of compounds by their structural properties are powerful tools to design highly active molecules. In this sense, quantitative structure–activity relationship (QSAR) studies have been successfully applied for modeling biological activities of natural and synthetic chemicals [17]. QSAR studies have been carried out for modeling activities of several kinds of ARIs. Some recent reports have linked structural features of the ligands with their AR inhibition by using topology indexes [18,19], three-dimensional (3D)-QSAR methodologies [20,21], artificial neural networks [21–24], etc.

The current study involves the development of 3D-QSAR models to predict and interpret the AR inhibitory activity of a set of 73 flavonoids [7] (the chemical structures are shown in Table 1). Stefanic-Petek et al. reported a QSAR model for describing these compounds by using multilinear regression analysis with classical and quantum chemical descriptors [7]. After that, Fernández et al. reported linear and nonlinear QSAR models, and found that struc-

* Tel.: +56 71 201 662; fax: +56 71 201 561.
E-mail addresses: jcaballero@utalca.cl, jmcr77@yahoo.com.

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.08.005

Table 1
Experimental and predicted AR inhibitory activities of flavonoids used in the current study.



Compound[a]	Scaffold	Substituents[b]	log 1/IC_{50}		
			Observed	Predicted	
				CoMFA[c]	CoMSIA[d]
1[e]	A	5,7,3′,4′-OH; 3,6-OCH_3	7.59	6.98	6.51
2	A	3′,4′-OH; 5,6,7,8-OCH_3	7.49	7.59	6.91
3[f]	A	6,3′,4′-OH; 5,7,8-OCH_3	7.44	7.16	6.89
4[e]	A	5,7,3′,4′-OH; 6-OCH_3; 8-CH_2Ph	7.47	7.23	7.46
5	A	5,3′,4′-OH; 6,7,8-OCH_3	7.41	**7.09**	**6.89**
6[e]	A	3′,4′-OH; 5,7,8-OCH_3	7.35	6.48	6.86
7	A	5,6,7,3′,4′-OH; 3-OCH_3	7.24	6.98	6.87
8[e]	A	5,6,3′,4′-OH; 7,8-OCH_3	7.19	7.15	6.77
9[f]	A	7,3′,4′-OH; 5,8-OCH_3	7.13	**6.71**	**7.11**
10	A	5,3′,4′-OH; 7,8-OCH_3	7.11	6.86	6.82
11[e]	A	3′,4′-OH; 5,6,7-OCH_3	7.04	6.75	7.00
12	A	5,6,7,3′,4′-OH; 8-OCH_3	6.92	7.18	6.95
13[f]	A	6,3′,4′-OH; 5,7-OCH_3	6.85	**6.58**	**6.87**
14[e]	A	4′-OH; 5,6,7,8-OCH_3	6.79	7.06	6.07
15	A	8,3′,4′-OH; 5,7-OCH_3	6.79	6.46	6.91
16[e]	A	3′,4′-OH; 3,5,7,8-OCH_3	6.77	6.28	6.89
17	A	5,6,7,3′,4′-OH	6.69	7.22	6.93
18[f]	A	5,3′,4′-OH; 6,7-OCH_3	6.92	**6.69**	**6.88**
19[e]	A	5,8,3′,4′-OH; 7-OCH_3	6.64	6.59	6.77
20	A	5,7,3′,4′-OH; 3,8-OCH_3	6.62	6.79	6.96
21[e]	A	6,4′-OH; 5,7,8-OCH_3	6.60	6.86	6.04
22	A	3′,4′-OH; 5,6,7-OCH_3	6.57	6.75	7.00
23[f]	A	5,7,3′,4′-OH; 8-OCH_3	6.55	**6.95**	**7.00**
24[e]	A	7,3′,4′-OH; 3,5,8-OCH_3	6.55	6.34	7.07
25	A	8-OCH_3; 5,6,7,3′,4′-$OCOCH_3$	6.52	6.73	6.29
26[e]	A	5,6,3′,4′-OH; 7-OCH_3	6.52	6.65	6.75
27[f]	A	6,3′,4′-OH; 3,5,7-OCH_3	6.52	**6.23**	**6.81**
28	A	5,3′,4′-OH; 3,6,7-OCH_3	6.46	6.32	6.85
29[e]	A	5,7,4′-OH; 6,8-OCH_3	6.39	6.88	6.19
30	A	5,4′-OH; 6,7,8-OCH_3	6.27	6.80	6.04
31	A	5,6,3′,4′-OH; 3,7-OCH_3	6.09	6.44	6.69
32 (quercetin)	A	3,5,7,3′,4′-OH	6.02	7.06	6.69
33[f]	A	5,6,4′-OH; 7,8-OCH_3	6.07	**6.87**	**5.92**
34	A	5,6,7,4′-OH; 8-OCH_3	5.92	5.94	5.83
35	A	5,6,7,4′-OH; 8,3′-OCH_3	5.92	5.56	5.43
36	A	5,4′-OH; 6,7-OCH_3	5.85	6.38	6.05
37 (quercitrin)	A	5,7,3′,4′-OH; 3-O-Rh	6.54	6.39	6.90
38[f]	B	2-COOH	5.66	**3.73**	**5.02**
39	B	2-$COOCH_2CH_3$	5.64	5.40	5.14
40	B	2-$COOCH_2$Ph	5.60	5.94	5.77
41	B	2-$COOCH(CH_3)_2$	5.51	5.84	5.44
42	A	5,7,4′-OH; 6,8,3′-OCH_3	5.35	5.62	5.15
43[f]	A	6,4′-OH; 5,7,8,3′-OCH_3	5.20	**4.58**	**5.29**
44	A	5,4′-OH; 6,7,3′-OCH_3	5.17	5.75	5.65
45	A	5,7-OH; 6,8,4′-OCH_3	5.14	4.97	4.89
46	A	5,6,7-OH; 8-OCH_3	5.09	5.21	5.14
47	A	5,6-OH; 7,8-OCH_3	5.08	4.44	4.74
48	A	3′,4′-OH; 5,6,7-OCH_3; 3-$COCH_3$	5.05	5.00	5.10
49[f]	A	5,3′-OH; 6,7-OCH_3; 4′-O-Glc	5.09	**5.64**	**5.83**
50	C	4-OH	4.84	2.83	2.29
51	A	5-OH; 6,7,3′-OCH_3; 4′-O-Glc	4.81	4.74	5.01
52	A	5-OH; 6,7-OCH_3; 4′-O-Glc	4.60	4.48	5.05
54[f]	A	5,7-OH; 6,8,3′-OCH_3; 4′-O-Glc	4.33	**5.39**	**5.79**

Table 1 (*Continued*)

Compound[a]	Scaffold	Substituents[b]	log 1/IC_{50}		
			Observed	Predicted	
				CoMFA[c]	CoMSIA[d]
55	A	4′-OH; 5,6,7,8,3′-OCH_3	4.74	4.57	5.34
56	A	5,4′-OH; 6,8,3′-OCH_3; 7-O-Glc	4.15	4.05	3.38
57[f]	D	4-OH; 7-OCH_3; 3-Ph	3.92	**1.81**	**2.26**
58	A	5,7-OH; 6,8,3′,4′-OCH_3	4.67	4.34	4.98
59	C	3-Ph; 4-OH	3.06	3.28	2.49
60	D	3-OH; 6-OCH_3	2.67	2.32	1.99
61	C	3-CN	2.67	2.71	1.98
62	C	3-COOH	2.67	2.66	1.70
63	A	5,4′-OH; 6,7,8,3′-OCH_3	4.42	4.39	5.03
64[f]	D	3-OH	1.45	**1.13**	**2.02**
65	D	3,8-COOH; 5-OCH_3	1.45	1.49	2.52
66	D	4-OH; 3,7-OCH_3	0.44	0.60	1.92
67	D	7-OCH_3; 4-CH_3	-1.14	1.02	1.45
69	C	4-CH_3	-	-	-
70	C	3-CH_3; 4-OH	-	-	-
71	A	5,6,4′-OH; 7,8,3′-OCH_3	3.96	3.88	5.04
72	A	6-OH; 5,7,8-OCH_3	4.44	4.12	4.88
73[f]	E	5,5′-OH; 7,2′,4′-OCH_3	3.50	**3.98**	**3.33**
74	E	7-OH; 6-OCH_3	–	–	–
75	E	5,4′-OH; 7,2′,5′-OCH_3	–	–	–

[a] Compounds **1**, **3**, **14**, **18**, **21**, **25**, **28**–**30**, **33**–**37**, **42**–**52**, **54**–**67**, and **69**–**75** are from Ref. [8]; compounds **2**, **4**–**13**, **15**–**17**, **19**, **20**, **22**–**24**, **26**, **27**, **31** are from Ref. [28]; compounds **38**–**41** are from Ref. [29] and compound **32** is from Ref. [30].
[b] Rh, rhamnose; Glc, glucose.
[c] CoMFA test-set predictions are in boldface.
[d] CoMSIA test-set predictions are in boldface.
[e] Compounds used to generate pharmacophore model.
[f] Test-set compounds.

tural features related to the molecular topologies and charges are related to the inhibitory activity of these compounds. In the present paper, we reported comparative molecular field analysis (CoMFA) [25], and comparative molecular similarity indices análisis (CoMSIA) models [26], which may throw some light on the requirements of the substituents for the further development of more potent ARIs. In addition, we reported a pharmacophore model by using GALAHAD (Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Database) [27], which contains information from the most active molecules, and can be used as a 3D search query to probe databases for new active structures.

2. Materials and methods

2.1. Compounds and biological data

The molecular structures used in this study were collected by Stefanic-Petek et al. [7]. In a recent work, Mercader et al. [19] tried to use all the 75 molecules from that paper, but they found some important errors in the representation of some of the compounds. They analyzed the references containing the experimental data [8,28], and used 56 compounds from the original dataset. In the present work, we used 73 compounds from the paper of Stefanic-Petek et al. [7] (the chemical structures and experimental activities are shown in Table 1). The structures and the activities were corrected, according to the references containing the experimental data [8,28–30]. IC_{50} was not reported for several compounds in these references; instead of this, percentages of inhibition of AR at different concentrations of the enzyme were reported. For these compounds, log dose–response curves were constructed, and the concentration of inhibitor which was necessary for 50% inhibition of activity (IC_{50}) was estimated from the least-square regression line of the log dose–response curve [31]. Compounds **69**, **70**, **74** and **75** were reported as inactive, therefore, they were not considered for carrying out QSAR and pharmacophore models.

Due to the mistakes observed in previous works, the compounds are provided in mol2 format as supplementary material. For clarity, the numbers of the compounds were kept as in Stefanic-Petek et al. report [7]. The molecules were previously optimized by using a semi-empirical method (see protocol in Ref. [22]). The activities were collected as log(1/IC_{50}) values, where IC_{50} refers to the molar concentration of the compound required for 50% inhibition of rat lens AR activity.

2.2. CoMFA and CoMSIA models

CoMFA and CoMSIA models were obtained using the Sybyl 7.3 software of Tripos [32]. All the molecules were aligned by an atom-by-atom least-square fit. We used the chromene structure as a template. For the 3D-QSAR calculations, the compound set was randomly divided into a training set (55 compounds) and a test-set (14 compounds). This subdivision was performed in such away that both sets represent equally well the chemical and biological properties of the whole data set. The molecules of the training set were placed in a rectangular grid extended beyond 4 Å in each direction from the coordinates of each molecule. The interaction energies between a probe atom (a sp^3 hybridized carbon atom with +1 charge) and all compounds were computed at the surrounding points, using a volume-dependent lattice with 2.0 Å grid spacing. Then, standard Sybyl parameters were used for a partial least squares (PLS) analysis. The numbers of components in the PLS models were optimized by using Q^2 value, obtained from the leave-one-out (LOO) cross-validation procedure, with the SAMPLS [33] sampling method. The number of components was increased until additional components did not increase Q^2 by at least 5% per added component. The CoMFA models were generated by using steric and electrostatic probes with standard 30 kcal/mol cutoffs. In the CoMSIA analyses, similarity is expressed in terms of steric occupancy, electrostatic interactions, local hydrophobicity, and H-bond donor and acceptor properties, using a 0.3 attenuation factor.

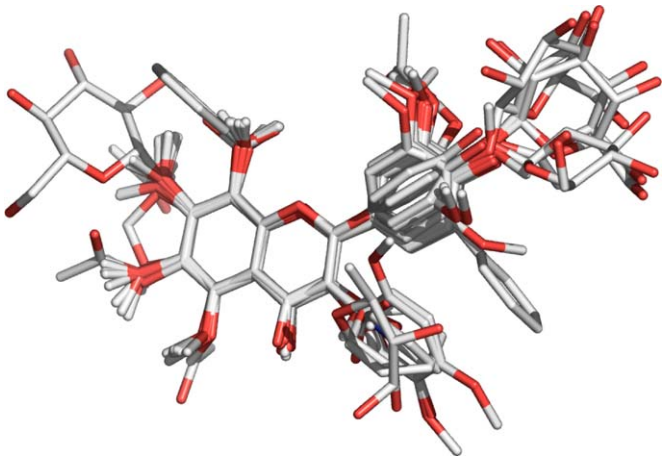


Fig. 1. Atom-by-atom superposition used for 3D-QSAR analysis.

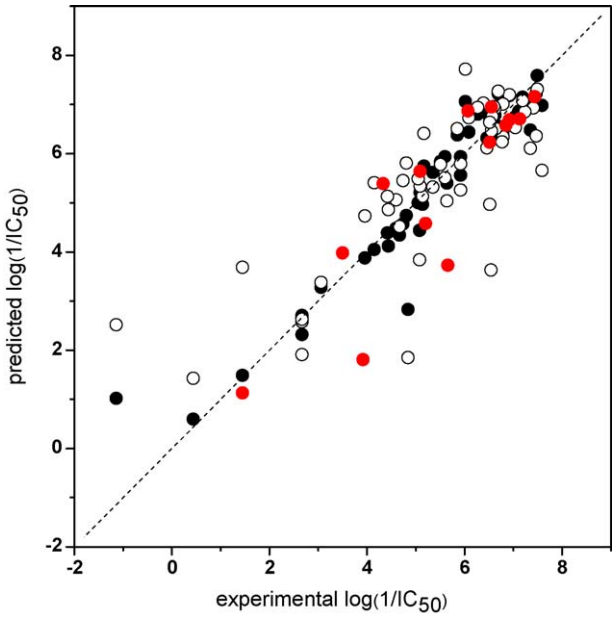


Fig. 2. Scatter plot of the experimental activities versus predicted activities for model CoMFA-SE: (●) training set predictions, (○) LOO cross-validated predictions, (●) test-set predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3. GALAHAD

Chemical function based pharmacophore models consist of a group of features that are located relative to each other in coordinate space as points surrounded by a sphere of tolerance. Each sphere represents the region in space that should be occupied by a certain chemical functionality capable of the kind of interaction specified by the feature type. GALAHAD is able to identify a set of ligand conformations that have an optimal combination of low strain energy (SE), steric overlap (SO), and pharmacophoric similarity (PhS). For the search of conformations, GALAHAD separates the model building process into two stages. First, a GA operates solely in the internal coordinates (torsional space), and then, a linear assignment routine aligns the conformers produced in cartesian space [34]. GALAHAD uses a true multi-objective (MO) function in which each term (SE, SO and PhS) is considered independently [35]. This MO functions are employed for three different purposes: to assess reproductive fitness, to select which candidates should survive to the next generation, and to rank models after Cartesian alignment of their constituent ligand conformers. The three MO functions constitute a multi-objective triage (MOTriage) approach, which make use of the Pareto rank for each individual model [36].

The ARIs were entered into SYBYL as mol2 files. Twelve of the most active compounds from the training set were selected to generate pharmacophore models. GALAHAD was run for 100 generations with a population size of 70 and a tournament pool size of 250. Default values were used for other settings. Only models with all the 12 ligands with contribution to the consensus feature were considered. Between the selected models, the one with the best SE, SO and PhS values based on Pareto ranking was selected as the best model.

3. Results and discussion

3.1. CoMFA and CoMSIA results

Fig. 1 shows the aligned molecules within the grid box (grid spacing 2.0 Å) used to generate the CoMFA and CoMSIA columns. The stepwise development of CoMFA and CoMSIA models using SAMPLS and different field combinations [37,38] is presented in Table 2. The predictability of the models is the most important criterion for assessment of both methods.

The best CoMFA model describing AR inhibition used both steric and electrostatic fields (model CoMFA-SE) and has a Q^2 value of 0.650 using four components; CoMFA models which include only one field showed less reliable statistics. CoMFA-SE model indicated similar contributions of both electrostatic and hydrophobic fields of 42.8 and 57.2%, respectively. The model explains 91.5 of the variance, has a low standard deviation ($s = 0.549$) and a high Fischer ratio ($F = 134.82$). The predictions of $\log(1/IC_{50})$ values for the 55 ARIs in the training set using CoMFA-SE models are shown in Table 1. The correlations between the calculated and experimental values of $\log(1/IC_{50})$ (from training and LOO cross-validation) are shown in Fig. 2. Model CoMFA-SE was also used to predict the inhibitory activities of the test-set compounds. We found that this model was able to describe the test-set variance with $R^2 = 0.781$ and standard deviation of test-set predictions (s_{test}) = 0.903. The test-set predicted values are listed in Table 1, and the correlations between the predictions and experimental values are represented in Fig. 2. This analysis revealed that the proposed model is able to predict successfully compounds that were not used in the training process.

The contour plots of the CoMFA steric and electrostatic fields are presented in Fig. 3 for the modeled AR inhibitory activities. In this figure, green and yellow isopleths indicate regions where bulky groups favored and disfavored the activity, respectively. A green isopleth is shown in Fig. 3 covering the ring at C2 of 4H-chromen-4-one scaffold of the compound **1**. This isopleth indicates that a phenyl ring at this position (scaffold A in Table 1) is mandatory for a highly AR inhibitory activity. A yellow isopleth extends in a more external part from the above-mentioned green isopleth. This suggests that substituents that are too large are not desired in this zone. For instance, compounds **49**, **51**, **52**, and **54** have an O-glucosyl group at position 4′ of scaffold A, and have less activity than analogous compounds containing hydroxyl group at this position (compounds **18**, **44**, **36**, and **42**, respectively). Other yellow isopleth is located near the position 3 of the compound **1** (Fig. 3). This suggests that substituents at this position are not favorable for the inhibitory activity. It is noteworthy that the poorly active compounds **73**, **74**, and **75** contain substituents at this position. Other yellow isopleth between C6 and C7 indicate that bulky substituents are not allowed at this position. In fact, when OH group at C7 of compound **42** is changed by an O-Glc (compound **56**), the activity decreases. In addition, the presence of a dioxole group (scaffold C in Table 1) does not contribute to a highly activity. Finally, yel-

Table 2
Stepwise development of CoMFA and CoMSIA models by using SAMPLS and different field combinations.[a]

	NC	Q^2	s_{CV}	Fields included in the model				
				Steric	Electrostatic	Hydrophobic	H-bond donor	H-bond acceptor
CoMFA-S	4	0.626	1.153	X				
CoMFA-E	4	0.590	1.206		X			
CoMFA-SE	**4**	**0.650**	**1.114**	**X**	**X**			
CoMSIA-S	3	0.623	1.146	X				
CoMSIA-E	6	0.521	1.267		X			
CoMSIA-H	3	0.623	1.145			X		
CoMSIA-D	3	0.501	1.317				X	
CoMSIA-A	3	0.559	1.239					X
CoMSIA-SE	5	0.646	1.132	X	X			
CoMSIA-SH	3	0.644	1.112	X		X		
CoMSIA-SD	4	0.631	1.145	X			X	
CoMSIA-SA	3	0.645	1.111	X				X
CoMSIA-ED	1	0.519	1.269		X		X	
CoMSIA-EA	3	0.554	1.246		X			X
CoMSIA-HD	3	0.676	1.061			X	X	
CoMSIA-HA	3	0.692	1.036			X		X
CoMSIA-DA	4	0.539	1.278				X	X
CoMSIA-SED	3	0.568	1.226	X	X		X	
CoMSIA-SEA	3	0.603	1.176	X	X			X
CoMSIA-SHD	3	0.687	1.044	X		X	X	
CoMSIA-SHA	**3**	**0.693**	**1.033**	**X**		**X**		**X**
CoMSIA-SDA	4	0.610	1.177	X			X	X
CoMSIA-HDA	3	0.679	1.057			X	X	X
CoMSIA-ALL	3	0.663	1.082	X	X	X	X	X

[a] NC is the number of components from PLS analysis, Q^2 and s_{CV} are the correlation coefficient and standard deviation of the leave-one-out (LOO) cross-validation, respectively. The best CoMFA and CoMSIA models are indicated in boldface.

low isopleth near C5 indicate that very large substituents are not allowed at this position. In Fig. 3, blue isopleths indicate regions where an increase of positive charge enhances the activity, and red isopleths indicate regions where more negative charges are favorable for activity. Red isopleths near C3′ and C4′ of compound **1** (see Table 1) indicates that OH groups at these positions are important for the AR inhibitory activity. Blue and red isopleths surrounding C4, C5, C6, C7 and C8 indicate that groups containing electronegative atoms, such as OCH_3 and OH, should occupy these positions.

In comparison to CoMFA, CoMSIA methodology has the advantage of exploring more fields. A more statistically robust model was obtained from the CoMSIA study. The best CoMSIA model included steric, hydrophobic and H-bond (HB) acceptor fields (CoMSIA-SHA) and has a Q^2 value of 0.693 using three components. CoMSIA-SHA model indicated contributions of steric, hydrophobic and HB acceptor fields of 11.2, 42.6 and 46.2%, respectively, explains 84.2 of the variance, has a low standard deviation ($s = 0.741$) and a high Fischer ratio ($F = 90.81$). CoMSIA models which include only one field showed less reliable statistics, and the addition of other fields did not produce an improvement in the internal validation. The predictions of $\log(1/IC_{50})$ values for the 55 ARIs in the training set using CoMSIA-SHA model are shown in Table 1. The correlations between the calculated and experimental values of log(1/IC50) (from training and LOO cross-validation) are shown in Fig. 4. Model CoMSIA-SHA was also used to predict the inhibitory activities of the test-set compounds. We found that this model was able to describe the test-set variance with $R^2 = 0.833$, and $s_{test} = 0.756$. The test-set predicted values are listed in Table 1, and the correlations between the predictions and experimental values are represented in Fig. 4. This analysis revealed that the proposed model is able to predict successfully compounds that were not used in the training process.

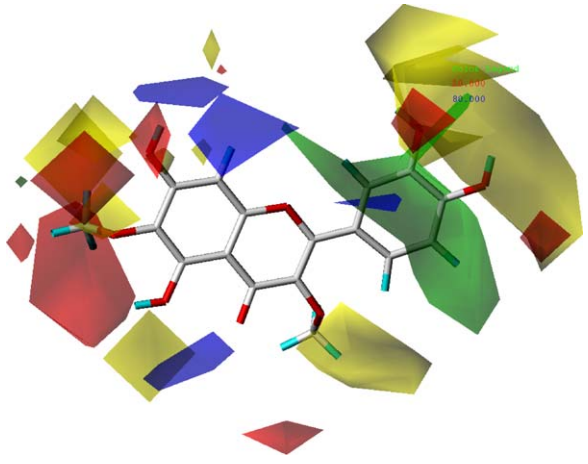


Fig. 3. CoMFA contour maps for ARIs (CoMFA-SE model). Compound **1** is shown inside the field. Green isopleths indicate regions where bulky groups enhance the activity, and yellow isopleths indicate regions where bulky groups disfavor the activity. Blue isopleths indicate regions where an increase of positive charge enhances the activity, and red isopleths indicate regions where more negative charges are favorable for activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The contour plots of the CoMSIA steric, hydrophobic and HB acceptor fields are presented in Fig. 5. For simplicity, the interaction between only the most active compound and the contour map is shown. In Fig. 5a, green isopleth indicates region where bulky groups favored the activity, respectively. As in contours of CoMFA-SE model, the contour plots of CoMSIA-SHA model showed a green isopleth covering the ring at C2 of 4H-chromen-4-one scaffold of the compound **1**; it was interpreted in the same manner as in the above-mentioned CoMFA model. In Fig. 5b, yellow isopleths indicate regions where hydrophobic/hydrophilic groups enhance/decrease the activity, and white isopleths indicate regions where hydrophobic/hydrophilic groups decrease/enhance the activity. A yellow isopleth covering the aromatic ring at C2 of 4H-chromen-4-one scaffold of compound **1** indicate that hydrophobic groups are allowed in these regions. Instead, the presence of white isopleths at the other side of the molecule, near C4, C5, C6, and C7, suggests that hydrophilic groups are necessary at these positions. In Fig. 5c, magenta isopleths represent areas where HB acceptor groups favor the activity. A big magenta isopleth appears near C3′ and C4′ of the ring at C2 of 4H-chromen-4-one scaffold.

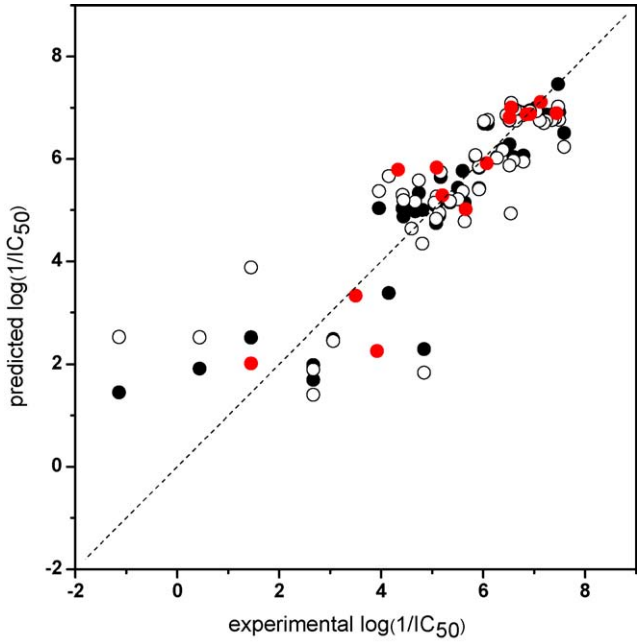


Fig. 4. Scatter plot of the experimental activities versus predicted activities for model CoMSIA-SHA: (●) training set predictions, (○) LOO cross-validated predictions, (●) test-set predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This result is consistent with previous results of Okuda et al. [8], who found that substituents at C3′ and C4′ position of the phenyl ring should have electron-donating properties. Finally, other small magenta isopleth near C7 suggests that HB acceptor groups at these zones favor the activity.

3.2. Pharmacophore results

GALAHAD models were derived by using 12 active ligands as a training set (these 12 compounds are shown in Table 1). Thirty pharmacophore models were retained after GALAHAD run. Each of the obtained models represents a different tradeoff among the conflicting demands of maximizing steric consensus, maximizing pharmacophore consensus, and minimizing energy. All the obtained models were derived from more than nine ligands. In addition, they had Pareto rank 0; this means no one model is superior to any other. One model had a very high energy (SE = 1.8×10^8), leading to its exclusion from our analysis. During GALAHAD runs, it is recognized that high SE values are due to steric clashes [39]. The algorithm retains these models to keep good characteristics to be passed on to less strained offspring during genetic algorithm (GA) process.

Only the GALAHAD models derived from all the 12 ligands of the training set were compared according to Pareto ranking. Table 3 shows SE, SO and PhS values for models with all the 12 ligands with contribution to the consensus feature. Minimum and maximum values for each characteristic between all the obtained 30 models are also reported in this table. Small value of SE and high values of SO and PhS are desired for the best model. After excluding the high value SE = 1.8×10^8, the higher value between all the models (the new maximum) is 83.95; the models containing all the 12 ligands had values between 13.88 (the minimum) and 83.95 (the maximum), in this sense SE value varies widely between the considered models. SO had a small variation between the minimum (SO = 1893.9) and the maximum (SO = 2161.4) considering all the 30 models. For models derived from all the 12 ligands, SO takes values between 1893.9 and 2034.0. In general, SO values were closer to the minimum for these models; this means that the steric overlap between the ligands was sacrificed to include all the compounds of the training set. Finally, PhS had a small variation between 354.0 (the minimum) and 378.0 (the maximum). For models derived from all the 12 ligands, PhS takes values between 354.4 and 378.0; according to this, PhS values do not have a preferential distribution for these models.

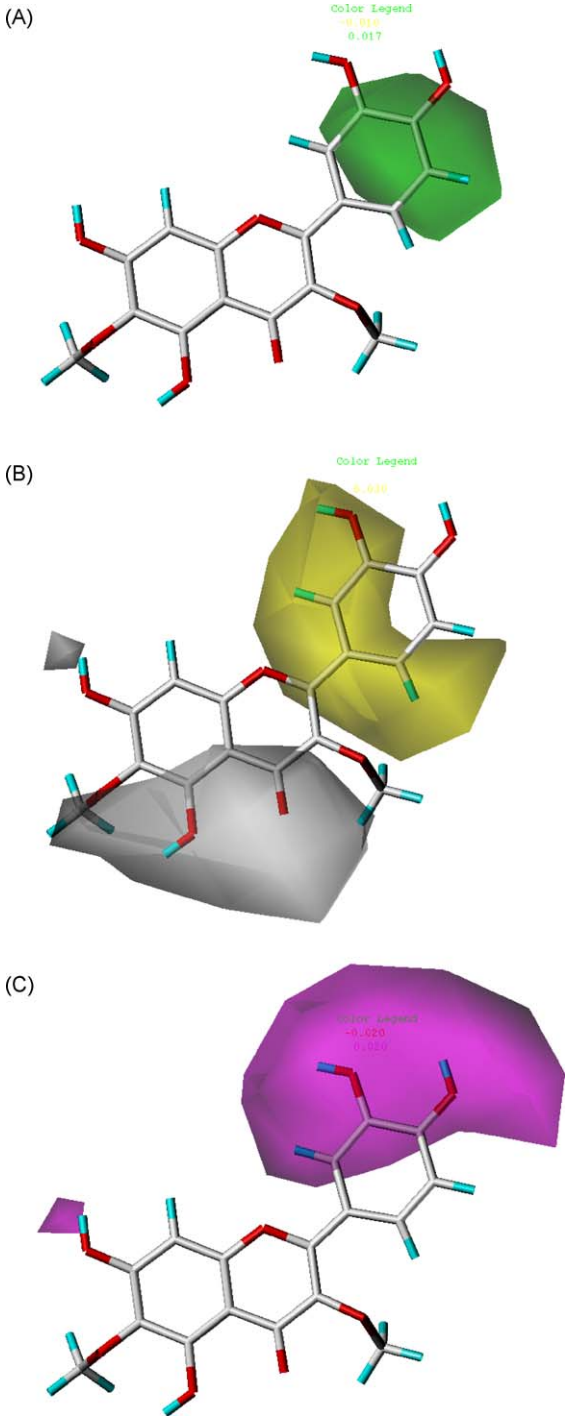


Fig. 5. CoMSIA contour maps for ARIs (CoMSIA-SHA model). Compound **1** is shown inside the field. (a) Steric field: green isopleths indicates region where bulky groups enhance the activity. (b) Hydrophobic field: Yellow and white isopleths indicate regions where hydrophobic/hydrophilic groups enhance/decrease the activity, and white isopleths indicate regions where hydrophobic/hydrophilic groups decrease/enhance the activity. (c) HB acceptor field: Magenta represents areas where HB acceptors favor the activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Strain energy (SE), steric overlap (SO) and pharmacophoric similarity (PhS) values for GALAHAD models with all the 12 ligands with contribution to the consensus feature.

Model	Features	SE	SO	PhS
G_2	11	13.88	1893.9	369.3
G_4	10	18.10	1894.2	370.0
G_5	11	83.95	1991.3	378.0
G_7	**9**	**29.13**	**1990.7**	**377.7**
G_11	11	22.59	1924.3	369.6
G_14	11	21.20	1995.3	354.4
G_16	9	20.97	1909.2	370.4
G_18	11	62.32	2034.0	354.4
G_19	11	23.89	1942.9	359.4
G_26	11	27.55	1929.7	359.8
Min[a]		13.88	1893.9	354.0
Max[a]		1.80×10^8	2161.4	378.0

The selected model (model G_7) is indicated in boldface.
[a] Minimum and maximum values between all the obtained 30 models.

With the intention to select the best model, we constructed a 3D plot to visualize the Pareto surface (Fig. 6). Considering only the SE and SO criteria, the best of all models lies in the upper left hand corner of the graph in Fig. 6b, where the energy is low and the steric score is high. In terms of SE and PhS criteria, the best of all models lies in the upper left hand corner of the graph in Fig. 6c, where the energy is low and the HB score is high. Finally, in terms of PhS and SO scores, the best of all models now lies at the upper-right corner, where both are high (Fig. 6d). In Fig. 6, the ideal model is represented with a blue circle. Between the considered models, model G_7 (represented with a red circle in Fig. 6) has the better position since fulfills the three criteria and it is one of the closest points to the blue circle. Other models that could have been chosen are shown in Fig. 6: models G_5, G_14 and G_18. Model G_5 has high SO and PhS values, but it was ruled out by also presenting the higher SE value. Model G_14 has a high SO value and a small SE value, but it was ruled out by presenting a small PhS value. Model G_18 has the higher SO value, but it presents a high SE value and a small PhS value.

The best GALAHAD model G_7 is displayed in Fig. 7. It is comprised of 12 substructures, one conformer for each molecule

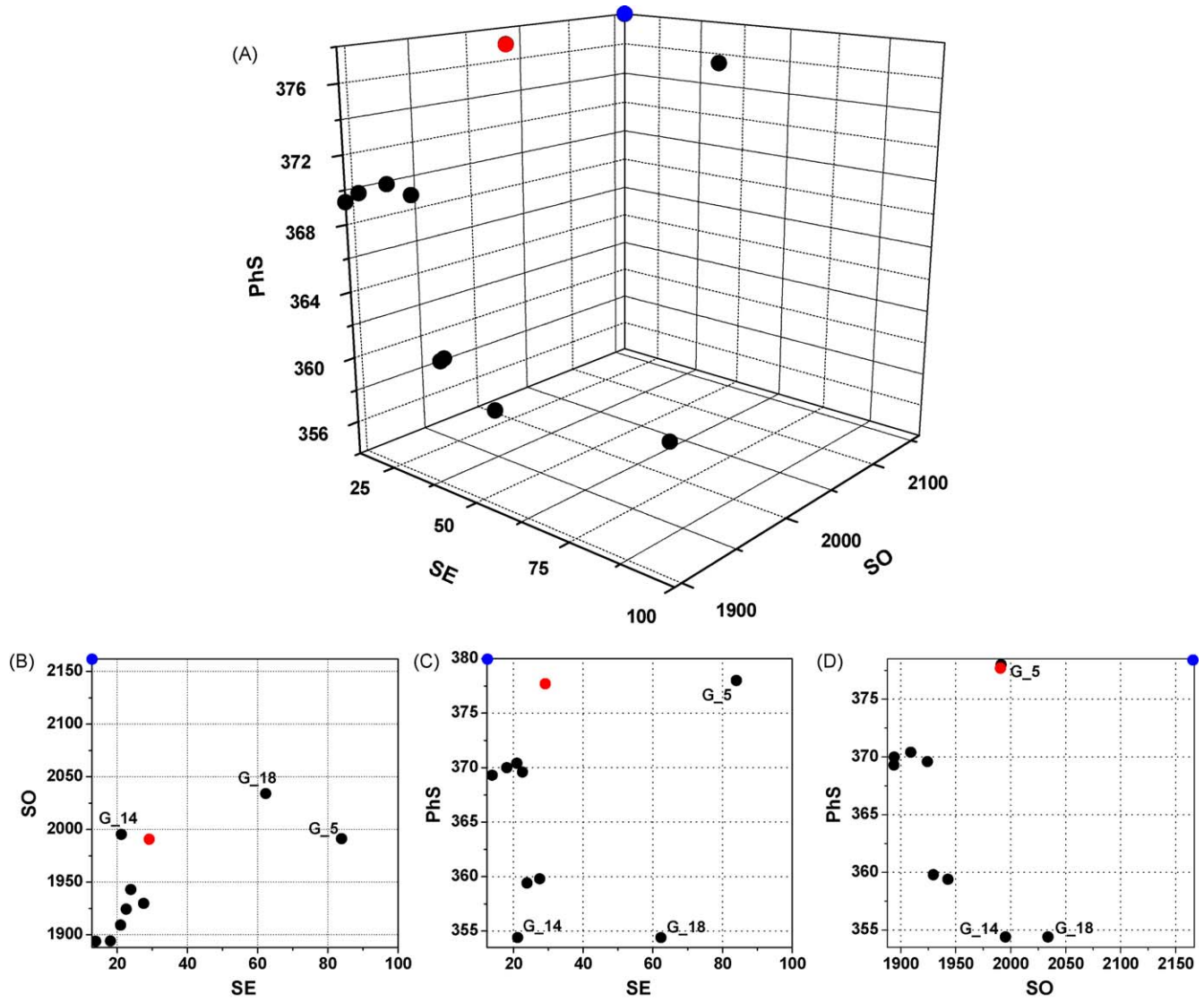


Fig. 6. Plot of the strain energy (SE), steric overlap (SO) and pharmacophoric similarity (PhS) values for GALAHAD models with all the 12 ligands with contribution to the consensus feature. (a) 3D plot; (b) plot of SO vs. SE; (c) plot of PhS vs. SE; (d) plot of PhS vs. SO. Red circle represents model G_7, blue circle represents the ideal best scoring according to minimum and maximum values obtained between the all 30 models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

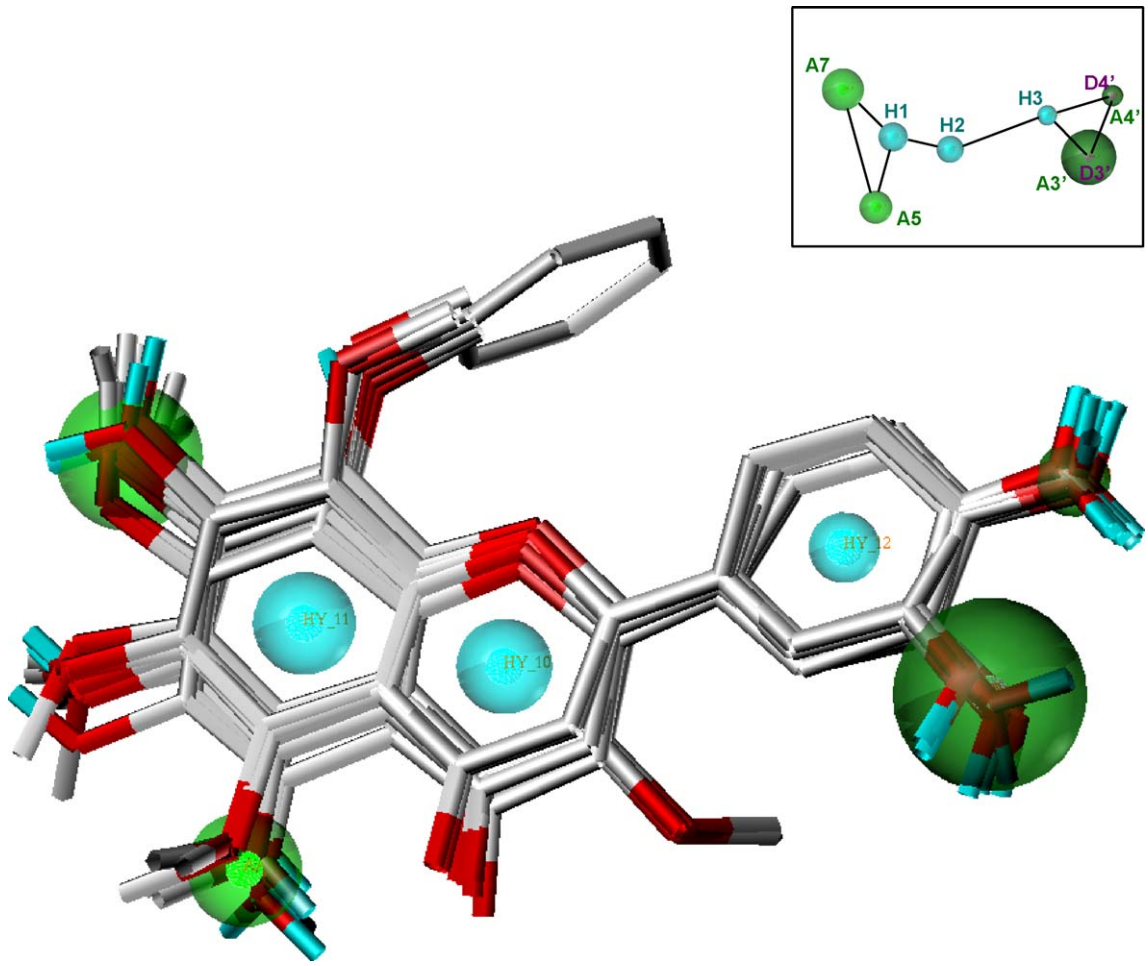


Fig. 7. Selected pharmacophore model G_7 and molecular alignment of the compounds used to elaborate the model. Cyan, green and magenta spheres are represented for hydrophobes, HB acceptors, and HB donors, respectively. Pharmacophore features are denoted in the upper-right corner. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in the training set, plus a 3D database query derived from features more or less shared among them. All conformers aligned represent low-energy conformations of the molecules, and it can be seen that the final alignment shows a satisfactory superimposition of the pharmacophoric points. In Fig. 7, cyan, green and magenta spheres indicate hydrophobes, HB acceptors, and HB donors, respectively. Model G_7 includes 9 pharmacophore features: three hydrophobes (H1, H2 and H3), four HB acceptors (A5, A7, A3′ and A4′), and two HB donors (D3′ and D4′). The three hydrophobic moieties of the pharmacophore schematically represented by H1, H2 and H3 reflect the need for a large hydrophobic structure as the skeleton of the ARIs, a requirement fulfilled by the flavonoids. The HB acceptor and donor moieties A3′/A4′ and D3′/D4′ are located at the oxygen atoms on C3′/C4′ of the ring at C2 of 4H-chromen-4-one scaffold. These features reflect the importance of OH groups at these positions of the flavonoids for AR inhibitory activity; this result is consistent with previous results of Okuda et al. [8]. A5 and A7 are located at the oxygen atoms on C5 and C7, respectively, of the 4H-chromen-4-one scaffold. These features reflect the importance of groups with HB acceptor atoms at these positions to have a good AR inhibitory activity.

We evaluated how well the model identifies active compounds. For this, model G_7 was used as a template in aligning the full dataset. We ran a partial least square (PLS) analysis by using the features of model G-7 as molecular descriptors to predict the activities of all compounds. The scatter plot of the calculated versus experimental values of $\log(1/IC_{50})$ using model G_7 are shown in Fig. 8. According to this graph, model G_7 was able to identify the active compounds from the set of molecules which were not used in deriving the model. 22 molecules that do not participate in the generation of model G_7, have $\log(1/IC_{50})$ values above 6 (Fig. 8). From them, 18 were predicted with an activity greater than 6. Under this criterion, 81.8% of the active molecules were correctly pre-

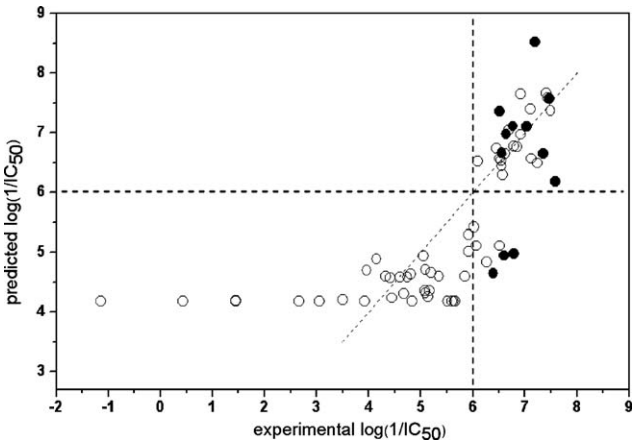


Fig. 8. Scatter plot of the experimental activities versus predicted activities for pharmacophore model G_7: (●) predictions for compounds used to elaborate the model, (○) predictions for the remaining compounds.

dicted. On the other hand, 35 molecules that do not participate in the generation of model G_7, have log(1/IC_{50}) values below 6; none was predicted with an activity greater than 6. This result indicates that model G_7 can be used as a theoretical screening tool able to discriminate between active and inactive molecules, and consequently, it can predict whether a new molecule inhibits AR.

4. Conclusions

Predictive 3D-QSAR and pharmacophore models were derived for flavonoids as inhibitors of AR, which should be useful for assisting the design of active compounds. Such models correlate well structural features with inhibitory activities against AR and bring valuable information about the relevant characteristics of inhibitors.

First, CoMFA and CoMSIA approaches were developed to derive structure–activity relationships. Reliable models were obtained by using steric and electrostatic CoMFA fields, and by using steric, hydrophobic and HB acceptor CoMSIA fields. Moreover, contour plots may help identify relevant regions where any change can affect binding preference. Furthermore, they may be helpful in identifying important features contributing to interactions between the studied flavonoids and the active site of AR. According to the obtained statistics, prediction of AR activities with sufficient accuracy should be possible by using these models.

Afterwards, pharmacophore models were derived by using GALAHAD. Pharmacophore models have often proved to be useful tools for rationalizing ligand–target interaction and for making this information available to virtual screening techniques. A model derived from 12 active compounds was obtained. The model includes 9 pharmacophore features: three hydrophobes, four HB acceptors, and two HB donors. The obtained GALAHAD queries comprise flexible search queries that can be used as a search engine in the large compounds database.

Acknowledgement

J.C. thanks “Becas Universidad de Talca” for financial support through a doctoral fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.08.005.

References

[1] M. Lorenzi, The polyol pathway as a mechanism for diabetic retinopathy: attractive, elusive, and resilient, Exp. Diabetes Res. 2007 (2007) 61038.
[2] S.S. Chung, E.C. Ho, K.S. Lam, S.K. Chung, Contribution of polyol pathway to diabetes-induced oxidative stress, J. Am. Soc. Nephrol. 14 (2003) S233–236.
[3] Z. Kyselova, M. Stefek, V. Bauer, Pharmacological prevention of diabetic cataract, J. Diabetes Complications 18 (2004) 129–140.
[4] Y. Hamada, J. Nakamura, Clinical potential of aldose reductase inhibitors in diabetic neuropathy, Treat. Endocrinol. 3 (2004) 245–255.
[5] J.M. Forbes, M.T. Coughlan, M.E. Cooper, Oxidative stress as a major culprit in kidney disease in diabetes, Diabetes 57 (2008) 1446–1454.
[6] S. Miyamoto, Recent advances in aldose reductase inhibitors: potential agents for the treatment of diabetic complications, Expert Opin. Ther. Patents 12 (2002) 621–631.
[7] A. Stefanic-Petek, A. Krbavcic, T. Solmajer, QSAR of flavonoids. 4. Differential inhibition of aldose reductase and p56lck protein tyrosine kinase, Croat. Chem. Acta 75 (2002) 517–529.
[8] J. Okuda, I. Miwa, K. Inagaki, T. Horie, M. Nakayama, Inhibition of aldose reductases from rat and bovine lenses by flavonoids, Biochem. Pharmacol. 31 (1982) 3807–3822.
[9] H. Matsuda, T. Morikawa, I. Toguchida, M. Yoshikawa, Structural requirements of flavonoids and related compounds for aldose reductase inhibitory activity, Chem. Pharm. Bull. 50 (2002) 788–795.
[10] C. Koukoulitsa, F. Bailly, K. Pegklidou, V.J. Demopoulos, P. Cotelle, Evaluation of aldose reductase inhibition and docking studies of 6′-nitro and 6′,6″-dinitrorosmarinic acids, Eur. J. Med. Chem. 45 (2010) 1663–1666.
[11] G. Bruno, L. Costantino, C. Curinga, R. Maccari, F. Monforte, F. Nicolò, et al., Synthesis and aldose reductase inhibitory activity of 5-arylidene-2,4-thiazolidinediones, Bioorg. Med. Chem. 10 (2002) 1077–1084.
[12] N. Hotta, Y. Akanuma, R. Kawamori, K. Matsuoka, Y. Oka, M. Shichiri, et al., Long-term clinical effects of epalrestat, an aldose reductase inhibitor, on diabetic peripheral neuropathy, Diabetes Care 29 (2006) 1538–1544.
[13] P Inskeep, R. Ronfeld, M. Peterson, N. Gerber, Pharmacokinetics of the aldose reductase inhibitor, zopolrestat, in humans, J. Clin. Pharmacol. 34 (1994) 760–766.
[14] A.M. Gonzalez, M. Sochor, P. McLean, The effect of an aldose reductase inhibitor (Sorbinil) on the level of metabolites in lenses of diabetic rats, Diabetes 32 (1983) 482–485.
[15] E.H. Akamine, T.C. Hohman, D. Nigro, M.H.C. Carvalho, R.D.C. Tostes, Z.B. Fortes, Minalrestatan aldose reductase inhibitor. Corrects the impaired microvascular reactivity in diabetes, J. Pharmacol. Exp. Ther. 304 (2003) 1236–1242.
[16] Y. Tanaka, T. Sawamoto, A. Suzuki, T. Kimura, Pharmacokinetics of zenarestat, an aldose reductase inhibitor, in male and female diabetic rats, Drug Metabol. Dispos. 21 (1993) 677–681.
[17] H. Kubinyi, QSAR: hansch analysis and related approaches, In: Methods and Principles in Medicinal Chemistry, R. Mannhold, P. Krokgsgaard-Larsen, H. Timmerman (Eds.), Weinheim, VCH, 1993.
[18] Y.S. Prabhakar, M.K. Gupta, N. Roy, Y. Venkateswarlu, A high dimensional qsar study on the aldose reductase inhibitory activity of some flavones: topological descriptors in modeling the activity, J. Chem. Inf. Model. 46 (2006) 86–92.
[19] A.G. Mercader, P.R. Duchowicz, F.M. Fernández, E.A. Castro, D.O. Bennardi, J.C. Autino, et al., QSAR prediction of inhibition of aldose reductase for flavonoids, Bioorg. Med. Chem. 16 (2008) 7470–7476.
[20] H. Liu, S. Liu, L. Qin, L. Mo, CoMFA, CoMSIA analysis of 2,4-thiazolidinediones derivatives as aldose reductase inhibitors, J. Mol. Model. 15 (2009) 837–845.
[21] S. Thareja, S. Aggarwal, T.R. Bhardwaj, M. Kumar, 3D-QSAR studies on a series of 5-arylidine-2,4-thiazolidinediones as aldose reductase inhibitors: a self-organizing molecular field analysis approach, Med. Chem. 6 (2010) 30–36.
[22] M. Fernández, J. Caballero, A.M. Helguera, E.A. Castro, M.P. González, Quantitative structure–activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds, Bioorg. Med. Chem. 13 (2005) 3269–3277.
[23] J.C. Patra, O. Singh, Artificial neural networks-based approach to design ARIs using QSAR for diabetes mellitus, J. Comput. Chem. 30 (2009) 2494–2508.
[24] L. Hu, G. Chen, R.M. Chau, A neural networks-based drug discovery approach and its application for designing aldose reductase inhibitors, J. Mol. Graph. Model. 24 (2006) 244–253.
[25] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[26] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4146.
[27] N.J. Richmond, C.A. Abrams, P.R.N. Wolohan, E. Abrahamian, P. Willett, R.D. Clark, GALAHAD. 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D, J. Comput. Aided Mol. Des. 20 (2006) 567–587.
[28] J. Okuda, I. Miwa, K. Inagaki, T. Horie, M. Nakayama, Inhibition of aldose reductase by 3′,4′-dihydroxyflavones, Chem. Pharm. Bull. 32 (1984) 767–772.
[29] P.F. Kador, J.H. Kinoshita, N.E. Sharpless, Aldose reductase inhibitors: a potential new class of agents for the pharmacological control of certain diabetic complications, J. Med. Chem. 28 (1985) 841–849.
[30] S.D. Varma, J.H. Kinoshita, Inhibition of lens aldose reductase by flavonoids – their possible role in the prevention of diabetic cataracts, Biochem. Pharmacol. 25 (1976) 2505–2513.
[31] P.F. Kador, J.H. Kinoshita, W.H. Tung, L.T. Chylack, Differences in the susceptibility of various aldose reductases to inhibition. II, Invest. Ophthalmol. Vis. Sci. 19 (1980) 980–982.
[32] SYBYL, version 7.3, 1699 South Hanley Rd., St. Louis, MO 63144, USA, Tripos Inc., SD.
[33] B.L. Bush, R.B. Nachbar, Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA, J. Comput. Aided Mol. Des. 7 (1993) 587–619.
[34] N.J. Richmond, P. Willett, R.D. Clark, Alignment of three-dimensional molecules using an image recognition algorithm, J. Mol. Graph. Model. 23 (2004) 199–209.
[35] V.J. Gillet, W. Khatib, P. Willett, P.J. Fleming, D.V.S. Green, Combinatorial library design using a multiobjective genetic algorithm, J. Chem. Inf. Comput. Sci. 42 (2002) 375–385.
[36] R. Clark, E. Abrahamian, Using a staged multi-objective optimization approach to find selective pharmacophore models, J. Comput. Aided Mol. Des. 23 (2009) 765–771.
[37] J. Caballero, M. Saavedra, M. Fernández, F.D. González-Nilo, Quantitative structure–activity relationship of rubiscolin analogues as δ opioid peptides using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), J. Agric. Food Chem. 55 (2007) 8101–8104.
[38] J. Caballero, M. Fernández, M. Saavedra, F.D. González-Nilo, 2D autocorrelation, CoMFA, and CoMSIA modeling of protein tyrosine kinases' inhibition by substituted pyrido[2,3-d]pyrimidine derivatives, Bioorg. Med. Chem. 16 (2008) 810–821.
[39] R.J. Dorfman, K.M. Smith, B.B. Masek, R.D. Clark, A knowledge-based approach to generating diverse but energetically representative ensembles of ligand conformers, J. Comput. Aided Mol. Des. 22 (2007) 681–691.