J Mol Model (2009) 15:837–845
DOI 10.1007/s00894-008-0439-0

ORIGINAL PAPER

CoMFA and CoMSIA analysis of 2,4-thiazolidinediones derivatives as aldose reductase inhibitors

Hong-Yan Liu · Shu-Shen Liu · Li-Tang Qin · Ling-Yun Mo

Received: 29 August 2008 / Accepted: 22 November 2008 / Published online: 9 January 2009

Abstract Diabetes remains a life-threatening disease. The clinical profile of diabetic subjects is often worsened by the presence of several long-term complications, for example neuropathy, nephropathy, retinopathy, and cataract. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed on a series of 2,4-thiazolidinediones derivatives as aldose reductase (ALR2) inhibitors. Molecular ligand superimposition on a template structure was finished by the database alignment method. The 3D-QSAR models resulted from 44 molecules gave q^2 values of 0.773 and 0.817, r^2 values of 0.981 and 0.979 for CoMFA and CoMSIA, respectively. The contour maps from the models indicated that a large volume group next to the R-substituent will increase the ALR2 inhibitory activity. In fact, adding a $-CH_2COOH$ substituent at the R-position would generate a new compound with higher predicted activity.

Keywords CoMFA · CoMSIA · 2,4-Thiazolidinedione derivative · Aldose reductase inhibitor · 3D-QSAR

Introduction

Diabetes is increasing dramatically, with the continuous improvement in people's standard of living, changes in dietary patterns, together with a lessening of intense labor and increased stress. Diabetes mellitus has become a common disease in many countries around the world, currently affecting 246 million people worldwide and expected to affect 380 million by 2025. In 2007, the five countries with the largest numbers of people with diabetes were India (40.9 million), China (39.8 million), the United States (19.2 million), Russia (9.6 million), and Germany (7.4 million) [1–3].

Aldose reductase (ALR2, EC1.1.1.21) is an enzyme of primary importance in the development of degenerative complications of diabetes mellitus, through its ability to reduce excess glucose into sorbitol with concomitant conversion of NADPH to $NADP^+$ (Fig. 1) [4, 5]. Diabetes can cause many chronic complications such as neuropathy, retinopathy, nephropathy and cataract, and so on [6–12].

One of the causes of complications in diabetes is the abnormal osmotic pressure caused by hyperthyroidism and polyol metabolic activity. When glucose in cells becomes hyperglycemic, exceeding a certain range, aldose reductase (ALR2) will be activated, and in turn the polyol pathway, triggering the transformation process of glucose to sorbitol. Accumulation of sorbitol will increase osmotic pressure, causing tissue edema and matrix changes, which result in diabetes complications [13]. An effective strategy of preventing and improving diabetic complications is to inhibit the activity of the key enzyme in polyol pathway, ALR2. Thus, it is necessary to discover and search for new safe and effective aldose reductase inhibitors (ARI).

During the past few years, a large number of compounds were synthesized and evaluated as ARIs, such as flavonoids [14–16], spirosuccinimides [17, 18], 2,4-thiazolidinediones [19–23] (TZD), among which flavonoids and 2,4-thiazolidinedione derivatives proved to be potent. Furthermore, TZD are a new class of drugs for the treatment of type 2

H.-Y. Liu · S.-S. Liu · L.-Y. Mo
Department of Material and Chemical Engineering,
Guilin University of Technology,
541004 Guilin, People's Republic of China

S.-S. Liu (✉) · L.-T. Qin
Key Laboratory of Yangtze River Water Environment,
Ministry of Education, College of Environmental Science
and Technology, Tongji University,
200092 Shanghai, People's Republic of China
e-mail: ssliuhl@263.net

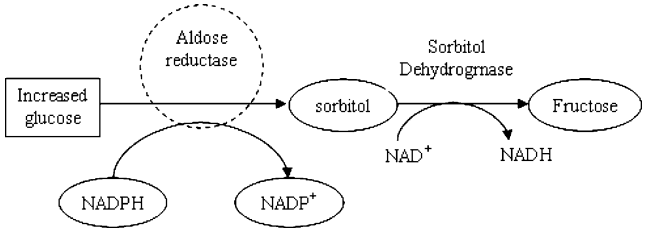


Fig. 1 Polyol pathway [6]

diabetes, and act by improving insulin sensitivity in adipose tissue, liver and skeletal muscle [24, 25]. This new type of ARI has been of great importance to researchers [19–23]. Based on the structures available, Rosanna Maccari and co-workers [19, 22, 23, 26] designed and synthesized a series of TZD as ARIs, and their inhibitory activities have also been measured (Table 1).

In the present study, we built the quantitative structure-activity relationships (QSAR) of the TZD derivatives using comparative molecular field analysis (CoMFA) [27] and comparative molecular similarity index analysis (CoMSIA) [28]. The purpose of this study is to offer some beneficial clues to structural modifications for designing new inhibitors with much higher inhibitory activities against ALR2, and to develop a predictive model for evaluating novel synthetic candidates. The result successfully demonstrated that QSAR is a useful tool for obtaining more effective inhibitor structures.

Materials and methods

Data sets

A data set of 55 compounds (structures and associated biological activities are given in Table 1) were taken from the literature [19–23]. Three skeleton structures (A, B, and C) listed in Table 1 are shown in Fig. 2. In vitro ARI activity values (IC_{50}) were converted into pIC_{50} according to the formula in Eq. 1. From Table 1, pIC_{50} values for 55 ARIs range from 4.10 to 6.89. Here, 17 ARIs display pIC_{50} values between 4.0 and 5.0, 18 ARIs between 5.0 and 6.0, and 19 ARIs between 6.0 and 7.0.

$$pIC_{50} = -\log IC_{50} \quad (1)$$

In order to validate and ensure the predictive potential of a model for the external ARIs, 55 ARIs was sorted ascending according to the pIC_{50} values, and 11 inhibitors (marked with "*" in Table 1) were equidistantly selected from Table 1 as an external test set; the remaining 44 compounds were used as a training set.

Molecular structure building and database alignment

All molecular modeling calculations were performed using SYBYL program package version 6.9 (SYBYL 6.9 Tripos Inc., http://www.tripos.com) on a Linux operating system [29]. Molecular building was done with a molecular sketch program. The molecular geometry of each compound was first minimized using a standard Tripos molecular mechanics force field with a 0.001 kcal/mol·Å energy gradient convergence criterion, and their charges were calculated by the Gasteriger-Hükel method [29]. Partial atomic charges were assigned to each atom and then energy minimization of each molecule was performed using the Powell method and Tripos standard force field with a distance-dependent dielectric function. The minimization was terminated when the energy gradient convergence criterion of 0.001 kcal/mol·Å was reached or when the 2,000-step minimization cycle limit was exceeded.

Molecular alignment is considered as one of the most sensitive parameters in CoMFA analysis [30, 31]. The quality and the predictive ability of the model are directly dependent on the alignment rule. Once the active conformation was determined, pharmacophore or common substructure alignment was carried out according to some rules. In this work, the superimposition of molecules was carried out by atom-based fitting of the heavy atoms of the ligands, shown in Fig. 3a. The compounds were fitted on the template molecule (compound **19**) making use of the heavy atoms of the common functionality present in all compounds of this series. The conformations of all aligned molecules of the training set are shown in Fig. 3b.

Comparative molecular field analysis

After alignment, CoMFA was used to study the QSAR of the inhibitors. The overlapped molecules were placed in a 3D lattice with regular grid spacing of 1 Å. Steric (Lennard-Jones potential) and electronic (Coloumb potentials) field energies were calculated using a sp^3 hybridized carbon atom as the steric probe atom and a +1 charge for the electrostatic probe with a cutoff energy of 30 kcal/mol. In order to speed up the analysis and reduce noise, column filtering was set at 2.0 kcal/mol.

Comparative molecular similarity index analysis

Five physicochemical properties, namely steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor, were evaluated. These fields were selected to cover the major contributions to ligand binding. Using all five CoMSIA descriptors for the explanatory variables, we performed a leave-one-out (LOO) run and a no validation partial least squares (PLS) analysis. Here, column filtering was set at

Table 1 The structures, observed pIC_{50} (Obs.), predicted pIC_{50} (Pred.) and their residuals (Res.) predicted by the CoMFA and CoMSIA models for the training set and test set

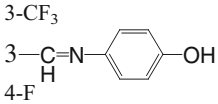
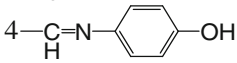
No.	Skeleton	Substituent		pIC_{50}	CoMFA		CoMSIA	
		R	R'	Obs.	Pred.	Res.	Pred.	Res.
1	A	H	3-F	5.04	5.04	0.00	4.93	0.11
2*	A	H	3-CH_3	5.03	4.76	0.27	4.93	0.1
3	A	H	3-OC_6H_5	5.21	5.40	-0.19	5.43	-0.22
4	A	H	3-OCH_3	4.88	4.93	-0.05	4.99	-0.11
5	A	H	3-CF_3	4.89	4.83	0.06	4.79	0.10
6	A	H		5.73	5.76	-0.03	5.81	-0.08
7	A	H	4-F	5.09	4.76	0.33	4.77	0.32
8	A	H	4-OCH_3	4.39	4.39	0.00	4.42	-0.03
9	A	H	4-CF_3	4.50	4.82	-0.32	4.97	-0.47
10*	A	H		4.64	4.72	-0.08	4.83	-0.19
11	A	CH_2COOCH_3	3-CH_3	5.00	4.89	0.11	5.08	-0.08
12	A	CH_2COOCH_3	3-OC_6H_5	5.88	5.65	0.23	5.58	0.30
13	A	CH_2COOCH_3	3-OCH_3	5.05	5.09	-0.04	5.13	-0.08
14	A	CH_2COOCH_3	3-CF_3	4.54	4.82	-0.28	4.92	-0.38
15*	A	CH_2COOCH_3	4-F	4.89	4.92	-0.03	4.91	-0.02
16	A	CH_2COOCH_3	4-CF_3	5.46	5.19	0.27	5.11	0.35
17*	A	CH_2COOH	3-F	6.13	6.44	-0.31	6.22	-0.09
18	A	CH_2COOH	3-CH_3	6.19	6.12	0.07	6.22	-0.03
19	A	CH_2COOH	3-OC_6H_5	6.89	6.85	0.04	6.72	0.17
20	A	CH_2COOH	3-OCH_3	6.32	6.31	0.01	6.27	0.05
21	A	CH_2COOH	3-CF_3	6.33	6.07	0.26	6.07	0.26
22	A	CH_2COOH	4-F	5.94	6.14	-0.20	6.06	-0.12
23	A	CH_2COOH	4-CF_3	6.34	6.40	-0.06	6.27	0.07
24	A	H	3-OH	4.97	5.10	-0.13	5.00	-0.03
25	A	H	4-OH	5.05	4.87	0.18	4.88	0.17
26	A	H	3-NH_2	4.69	4.69	0.00	4.49	0.20
27*	A	CH_2COOCH_3	4-OH	5.21	5.02	0.19	5.01	0.20
28	A	CH_2COOCH_3	3-NH_2	4.41	4.82	-0.41	4.63	-0.22
29	A	CH_2COOH	4-OC_6H_5	6.09	6.12	-0.03	6.28	-0.19
30*	A	CH_2COOH	4-$OCH_2C_6H_5$	6.55	6.84	-0.29	6.50	0.05
31	A	CH_2COOH	4-C_6H_5	6.59	6.53	0.06	6.55	0.04
32	A	CH_2COOH	3-OH	6.18	6.3	-0.12	6.22	-0.04
33*	A	CH_2COOH	4-OH	***6.82***	***6.25***	***0.57***	***6.15***	***0.67***
34	A	CH_2COOH	3-NO_2	6.31	6.25	0.06	6.32	-0.01
35*	A	CH_2COOH	4-OH	5.92	6.16	-0.24	***4.56***	***1.36***
36	A	H	3-OCH_3,4-OH	4.93	4.94	-0.01	4.92	0.01
37	A	CH_2COOH	3-OCH_3,4-OH	6.15	6.33	-0.18	6.18	-0.03
38*	A	CH_2COOH	3-OH, 4-OCH_3	6.25	6.29	-0.04	5.88	0.37
39	A	CH_2COOH	4-OCH_2COOH	6.22	6.18	0.04	6.24	-0.02
40	A	CH_2COOH	3-OCH_2COOH	6.64	6.63	0.01	6.75	-0.11
41	A	CH_2COOH	3-OCH_3,4-OCH_2COOH	6.59	6.57	0.02	6.58	0.01
42	A	CH_2COOH	3-OCH_2COOH,4-OCH_3	5.85	5.69	0.16	5.79	0.06
43	B	H	–	4.97	4.91	0.06	4.93	0.04
44	B	CH_2COOH	–	6.77	6.74	0.03	6.84	-0.07
45	C	H	3-OC_6H_5	4.10	4.04	0.06	4.16	-0.06
46*	C	H	4-OC_6H_5	4.39	4.09	0.30	4.15	0.24
47	C	H	4-$OCH_2C_6H_5$	4.50	4.49	0.01	4.51	-0.01
48	C	H	4-C_6H_5	4.18	4.1	0.08	4.14	0.04
49	C	CH_2COOCH_3	3-OCH_3	4.68	4.68	0.00	4.67	0.01
50	C	CH_2COOH	3-OC_6H_5	6.00	6.06	-0.06	5.96	0.04
51	C	CH_2COOH	4-OC_6H_5	5.55	5.55	0.00	5.51	0.04
52	C	CH_2COOH	4-C_6H_5	5.77	5.8	-0.03	5.76	0.01

Table 1 (continued)

No.	Skeleton	Substituent		pIC_{50}	CoMFA		CoMSIA	
		R	R'	Obs.	Pred.	Res.	Pred.	Res.
53	C	CH_2COOH	3-OCH_3	5.53	5.44	0.09	5.51	0.02
54	C	CH_2COOH	4-OCH_3	5.97	6.06	-0.09	6.04	-0.07
55*	C	CH_2COOH	4-OH	5.68	6.09	-0.41	6.04	-0.36

2.0 kcal/mol and a default value of 0.3 was used as the attenuation factor.

Statistic analysis

In CoMFA and CoMSIA calculations, the partial least squares (PLS) method may be a more viable application options. PLS was carried out with the LOO cross-validation procedure to determine the optimum number of components for the final non-cross-validated 3D-QSAR models. Firstly, the optimum number of components (N) used in the model derivation was chosen from the analysis with the highest cross-validated correlation coefficient (q^2) according to the formula Eq. 2.

$$q^2 = 1 - \frac{\sum (y_{obs} - y_{cal})}{\sum (y_{cal} - y_{mean})} \quad (2)$$

Where y_{obs} is the observed value, y_{cal} is the calculated value and y_{mean} is the average of all activity values.

The N is employed to perform no validation PLS analysis to obtain the final model parameters such as correlation coefficient r^2, standard deviation, S and F values. In order to speed up the analysis and reduce noise, column filtering was set at 2.0 kcal/mol. At the same time, CoMFA color contour maps were derived for the electrostatic and steric fields, and CoMSIA color contour maps were also derived for the electrostatic, steric, hydrophobic, hydrogen bond donor and acceptor fields.

Results and discussion

A data set of 55 ARIs was used to derive both the conventional CoMFA and CoMSIA models. In order to validate the predictive ability of the models, 44 ARIs were selected from all 55 ARIs to construct a training set and the remaining 11 ARIs formed a test set. Thus, two 3D-QSAR models were generated from training set molecules with two different optimal number of components (8 or 9). The cross-validated r^2 (q^2) values for the two models relating ALR2 inhibition are shown in Table 2.

CoMFA analysis

By using the setting CoMFA parameters, the LOO cross-validated correlation coefficient (q^2) of 0.733 was observed with an optimal numbers of components (N) of eight (Table 2). Then, internal non-cross-validated PLS regressions were computed using the previously obtained N giving regression coefficients r^2 of 0.965 and a standard error of estimate (S) of 0.162. The statistical parameters associated with the CoMFA model are listed in Table 2.

The predicted pIC_{50} values and residual values for training set and test set compounds are given in Table 1. Figure 4a shows the relationship between the CoMFA-predicted and observed pIC_{50} values of the non-cross-validated analyses for ARIs. From Fig. 4a, it can be seen that almost all points except No. 33 are located on the diagonal line.

CoMSIA analysis

CoMSIA is an extension of CoMFA methodology. CoMSIA is thought to be less affected by changes in molecular alignment and to provide more smooth and interpretable contour maps as a result of employing Gaussian type-distance dependence with the molecular similarity indices [21]. With five kinds of fields—steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor—applied in CoMSIA, a CoMSIA model was obtained. The statistical

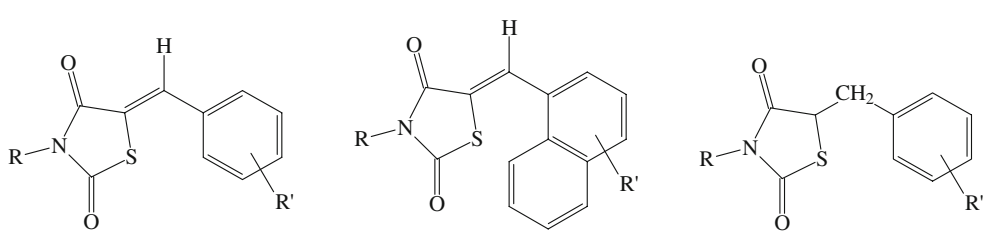


Fig. 2 Three skeleton structures (*A*, *B*, and *C*) of 55 ARI compounds

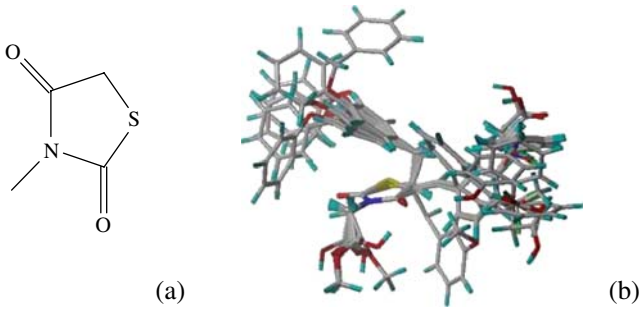


Fig. 3 Alignment of the training set: **a** heavy atoms in the common substructure for alignment, **b** alignment of compounds in the training set

details of CoMSIA model are listed in Table 2. Obviously, biological activity is concerned closely not only with the steric, electrostatic and hydrophobic properties, but also with H-bond donor and acceptor properties. The cross-validated correlation coefficient (q^2) and the conventional correlation coefficient (r^2) are 0.788 and 0.959, respectively, for CoMSIA. Final predicted/estimated versus observed pIC_{50} values for models and their residuals are given in Table 1. The relationship between pIC_{50} calculated by the CoMSIA model and observed values is shown in Fig. 4b, which shows that almost all points are close to the diagonal line except for Nos. 33 and 35.

Table 2 Statistical parameters of the CoMFA and CoMSIA models

QSAR parameters	CoMFA	CoMSIA
Training set		
q^2	0.733	0.788
r^2	0.965	0.959
N	8	9
S	0.162	0.179
F-value	120.493	87.749
Fraction of field contribution		
Steric	0.544	0.114
Electrostatic	0.456	0.220
Hydrophobic	–	0.204
Acceptor	–	0.316
Donor	–	0.146
Testing set		
q^2	0.657	0.694
r^2	0.952	0.991
N	4	5
S	0.236	0.148
F-value	120.493	87.749
r_0^2	0.999	0.934
k	0.995	1.036
r^2_{pred}	0.665	0.605

QSAR Quantitative structure-activity relationship, q^2 cross-validated correlation coefficient, r^2 conventional correlation coefficient, *N* optimal number of components, *S* standard error of estimation, *F* Fisher test value

Statistical parameters

In order to validate the predictive quality of the CoMFA and CoMSIA models, we calculated the statistical parameters of the test set with the formula based on references [32]–[36] (Table 2).

As can be seen from Table 2, the correlation coefficients of the test set are 0.952 and 0.991 for the CoMFA and CoMSIA models, respectively. The correlation coefficients for regressions through the origin (predicted vs observed activities), i.e., r_0^2, are 0.999 and 0.934, respectively. Moreover, the slope of regression lines of models is 0.995 and 1.036, which is close to 1. The predictive r^2 $\left(r^2_{pred}\right)$ [36–39] value based on molecules of the test set is 0.665 for the CoMFA model. The predictive r^2 value of the CoMSIA model is 0.605 (except for No. 35). The models are considered acceptable, because they satisfy all of the following conditions: (1) $q^2>0.5$, (2) $r^2>0.6$, (3) r_0^2 is close to r^2, such that $\left[\left(r^2 - r_0^2\right)/r^2\right] < 0.1$ and $0.85 <= k <=1.15$.

From above discussion, it can be concluded that both the CoMFA and the CoMSIA models have not only a good estimation ability but also a good predictive potential.

CoMFA and CoMSIA contour maps

Compared with two-dimensional QSAR methods, an important feature in the establishment of CoMFA and CoMSIA QSAR models is to provide the CoMFA steric, electrostatic fields and CoMSIA steric, electrostatic, H-acceptor, H-donor and hydrophobic fields contour maps. Such contour maps provide some information (such as steric, electrostatic, H-acceptor, H-donor, and hydrophobic) on factors affecting the activity of the study compounds. This is particularly important when increasing or reducing the activity of a compound by changing its molecular structure.

CoMFA and CoMSIA contour maps with a best-fit model were generated (Figs. 5, 6). The field energies at each lattice point were calculated as the scalar results of the coefficient, and the standard deviation associated with a particular column of the data table ("S.D.* coeff") was plotted as the percentage of the contribution to the CoMFA or CoMSIA equation. In the figures, the isocontour diagrams of the field contributions ("S.D.* coeff") for different properties calculated by the CoMFA and CoMSIA analysis are illustrated with exemplary ligands. Selectivity fields depict the change in binding preference occurring upon the change in molecular field around the ligands.

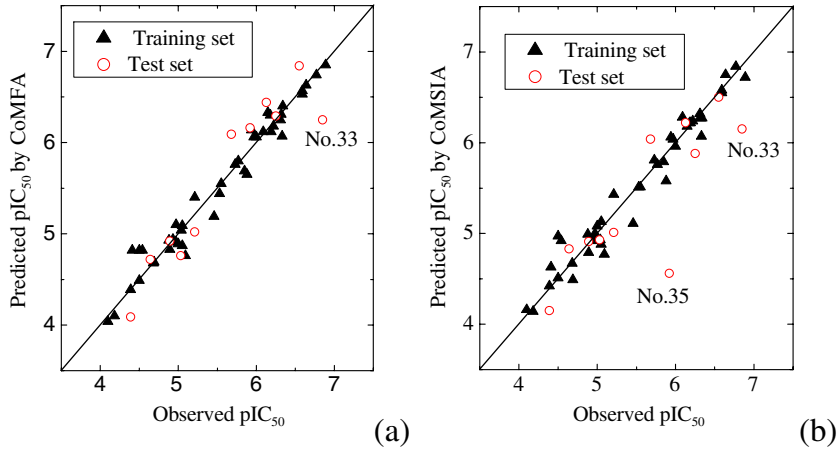


Fig. 4 Plot of the pIC50 values calculated vs observed using comparative molecular field analysis (CoMFA; **a**) and comparative molecular similarity analysis (CoMSIA; **b**)

Contour plots may help identify important regions where any change may affect binding preference. Furthermore, they may be helpful in identifying important features contributing to interactions between the ligand and the active site of a receptor. For the CoMFA steric, electrostatic fields and CoMSIA steric, electrostatic, acceptor, donor and hydrophobic fields, the contours represent 80% and 20% level contributions. For convenience, all similar contour map positions were labeled and shown in combination with the highly active compound No.19.

The CoMFA green contour indicates the area in which steric bulk might have a positive effect on activity while the yellow region is favorable for small groups. The blue contour indicates the region where positive groups are require for high activity while the red zone indicates a region favorable for negative groups (Fig. 5). For 55 ARIs, the effect of steric field is less important than the electrostatic field. The ratio of the contribution value of the two fields is 0.544/0.456.

As shown in Fig. 5a, substituents R (Table 1) linked to the big green contours indicate that steric bulk is favored for activity in these areas. This may be the reason why compounds with large CH_2COOH or CH_2COOCH_3 substituents in this area, e.g., compounds 12, 17, 18, 19, 27, 37, and 51 are higher in activity than molecules with small H substituents, such as compounds 1, 3, 4, 25, 36, 45, and 46, respectively. The yellow contour maps exist outwith the R positions, Fig. 5a. This can be explained by the fact that the compounds with a small CH_2COOH substitution in this area, e.g., compounds 22, 23, and 55, are highly active compared to compounds 15, 16, and 27 etc. which have CH_2COOCH_3 substituents in these areas leading to loss of activity. The CoMFA electrostatic contour plots are displayed in Fig. 5b. Two blue contour maps exist in the five-membered ring position, or above and below the phenyl ring, indicating that any positive charge or electron deficient substitute will enhance the activity at this position. For example, compound 23, which contains a CF_3 substituent, shows higher activity, and compound 22 with an F substituent exhibits lower activity. A large red contour map illustrates that negatively charged groups are favored in these areas. Accordingly, compounds 22, 23, and 55, which have CH_2COOH substitutions containing OH groups show good activity.

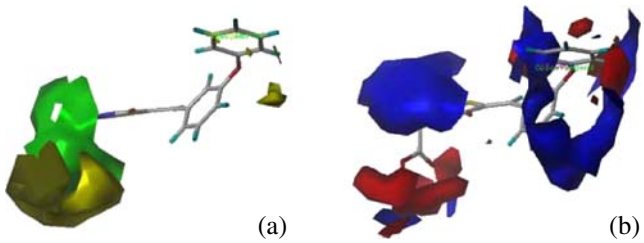


Fig. 5a,b CoMFA steric standard deviation (S.D.* coefficient) contour maps illustrating steric and electrostatic features in combination with compound No. 19. **a** *Green* contours show favorable bulky group substitution at that point while *yellow* regions show disfavorable bulky group for activity. **b** *Red* contours indicate negative charge favoring activity, whereas *blue* contours indicate positive charge favoring activity

Figure 6a and b describe the steric and electrostatic contour maps of the CoMSIA model. These contours are almost the same as the CoMFA-steric and electrostatic contours (Fig. 5). But in the CoMSIA model, two more contours appeared: a large green contour map on the phenyl ring and a large yellow contour map near this position.

Figure 6c shows the CoMSIA hydrogen bond acceptor field, denoted by magenta and red contours. Magenta contours represent regions where hydrogen bond acceptors on ligands are favorable, and red contours indicate regions where hydrogen bond acceptors on inhibitors are unfavorable for the activity. In Fig. 6c, the small magenta contour represents the five-membered ring, but the large red contours around it indicate that, in this position, any

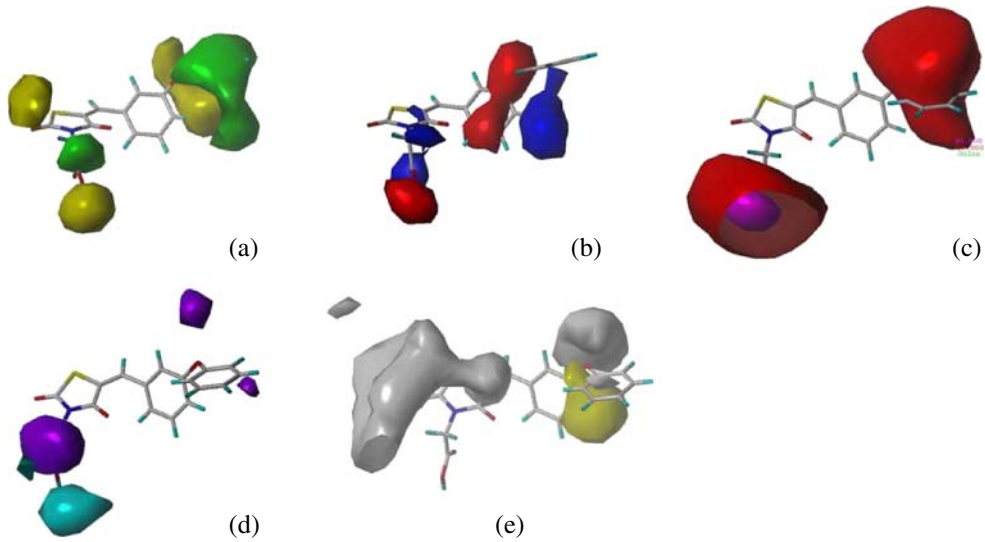


Fig. 6 CoMSIA S.D.* coefficient contour maps illustrating steric, electrostatic, acceptor, donor, hydrophobic features in combination with compound No. 19. **a** The *green* contour indicates a sterically favored region; *yellow* maps calls for a reduction of this potential to improve activity. **b** *Blue* indicates a positive charge preferred region to improve activity. **c** The *magenta* contour for H-bond acceptor group increases activity, *red* indicates the disfavored region. **d** The *cyan* contour for H-bond donor favored region, *purple* indicates the disfavored region. **e** The *yellow* contour for hydrophobic favored region, *white* indicates the hydrophilic favored region

substituent containing an acceptor group reduces the activity. For example, the carbonyl groups of compounds 20, 23, and 53 are more active than compounds 13, 16, and 49. The large red contours located at the R' position in Fig. 6c indicate that substituents with hydrogen-bond acceptors are unfavored in these areas. The -NH_2 group of the R' position of compounds 26 and 28 result in less activity.

The hydrogen-bond donor contours in Fig. 6d signify the regions of hydrogen-bond donor favorable (cyan) and unfavorable (purple) regions. One cyan contour is near the R adjacent to the five-membered ring, indicating that hydrogen bond donor functionalities in this region will enhance activity. Compounds 17, 20, 21, 32, 33, and 44 etc. are more active than compounds 1, 4, 5, 24, 25, and 26 etc., because they have an OH moiety located near this cyan contour. This cyan contour corresponds to ARIs, suggesting that this hydrogen bond donor group of ligands may form a strong hydrogen bond with the carbonyl oxygen of this residue and hence increase inhibitory potency. The cyan contour maps outside the R positions, signifying the position of donor groups present in the ARI compounds. Example compounds 11, 13, and 16 are less active than compounds 18, 20, and 23 because they have a CH_3 moiety located in this purple contour.

In Fig. 6e, the yellow and white contours enclose regions favorable for hydrophobic and hydrophilic groups, respectively. The white contour shown in Fig. 6e supports the importance of CH_2COOH substitutions at the R position that contains the -COOH moiety. The -COOH group in compounds 18–23 and compounds 29, 31, 32, 34, and 44 etc., satisfies the high activity condition. Another small yellow contour map located at the R' position, indicates that a hydrophobic function in this region will decrease activity, e.g., compounds 26 and 28, which contain an -NH_2 group, have lower activity.

Conclusions

The present study was aimed at deriving predictive models capable of elucidating the structural requirements for aldose reductase (ALR2) inhibitors. The 3D-QSAR analysis of TZD derivatives as ARIs was carried out using CoMFA and CoMSIA methods. A satisfactory pharmacophore model was obtained with LOO cross validation q^2 values of 0.733 and 0.788, conventional r^2 values of 0.965 and 0.959 for CoMFA and CoMSIA, respectively.

The CoMFA and CoMSIA models provided similar results and both exhibit dominant steric interactions. The relative contributions of steric fields are less important than those of electronic fields in both CoMFA and CoMSIA. The effects on activity of steric, electrostatic, hydrophobic, and H-bond donor and acceptor fields around the docked conformations were discussed in detail. Some implications can be drawn from this study to improve the activity and selectivity of ARIs, for example, a requirement for a CH_3COOH group at the R position to improve activity. Almost all compounds with CH_3COOH groups in this position have good activity.

The developed models not only possess promising predictive ability as shown by testing on the external test set, but should also be useful in elucidating the relationship between compound structures and biological activities. The models obtained will also serve as a basis for the design of novel ARI compounds with enhanced activity and other tailored properties.

Acknowledgments We are especially grateful to the National High Technology Research and Development Program of China (2007AA06Z417) and the National Natural Science Foundation of China (20577023, 20677056) and the Foundation for the Author of National Excellent Doctoral Dissertation of P. R. China (200355) for their financial support.

References

1. Zimmet P, Alberti KGMM, Shaw J (2001) Global and societal implications of the diabetes epidemic. Nature 414:782–787
2. Grubb T (2006) Diabetes Epidemic Out Of Control. http://diabetic-diet-secrets.com
3. Diamond J (2003) The double puzzle of diabetes. Nature 423:599–605
4. Koukoulitsa C, Zika C, Geromichalos GD et al (2006) Evaluation of aldose reductase inhibition and docking studies of some secondary metabolites, isolated from *Origanum vulgare* L. ssp. *hirtum*. Bioorg Med Chem 14:1653–1659
5. Brownlee M (2001) Biochemistry and molecular cell biology of diabetic complications. Nature 414:813–820
6. Ishii H, Tada H, Isogai S (1998) An aldose reductase inhibitor prevents glucose-induced increase in transforming growth factor-b and protein kinase C activity in cultured human mesangial cells. Diabetologia 41:362–364
7. Kasuya Y, Ito M, Nakamura J et al (1999) An aldose redutase inhibitor prevents the intimal thickening in coronary arteries of galactose-fed beagle dogs. Diabetologia 42:1404–1409
8. Calcutt NA, Freshwater JD, Mizisin AP (2004) Prevention of sensory disorders in diabetic Sprague-Dawley rats by aldose reductase inhibition or treatment with ciliary neurotrophic factor. Diabetologia 47:718–724
9. Cameron NE, Cotter MA, Basso M et al (1997) Comparison of the effects of inhibitors of aldose reductase and sorbitol dehydrogenase on neurovascular function, nerve conduction and tissue polyol pathway metabolites in streptozotocin-diabetic rats. Diabetologia 40:271–281
10. Petrash JM (2004) All in the family: aldose reductase and closely related aldoketo reductases. Cell Mol Life Sci 61:739–749
11. Hotta N (1995) New approaches for treatment in diabetes: aldose reductase inhibitors. Biomed Pharmacother 49:232–243
12. Settimo FD, Primofiore G, Motta CL et al (2005) Spirohydantoin derivatives of thiopyrano [2,3-*b*] pyridin-4(4*H*)-one as potent in vitro and in vivo aldose reductase inhibitors. Bioorg Med Chem 13:491–499
13. Gabbay KH (1973) The sorbitol pathway and the complications of diabetes. New Engl J Med 288:831–836
14. Fernández M, Caballero J, Helguera AM et al (2005) Quantitative structure-activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds. Bioorg Med Chem 13:3269–3277
15. Prabhakar YS, Gupta MK, Roy N et al (2006) A high dimensional QSAR study on the aldose reductase inhibitory activity of some flavones: topological descriptors in modeling the activity. J Chem Inf Model 46:86–92
16. Hu L, Chen G, Chau RMW (2006) A neural networks-based drug discovery approach and its application for designing aldose reductase inhibitors. J Mol Graph Model 24:244–253
17. Ko K, Won H, Won Y (2006) Quantitative structure-activity relationship of spirosuccinimide type aldose reductase inhibitors diminishing sorbitol accumulation in vivo. Bioorg Med Chem 14:3090–3097
18. Ko K, Won Y (2005) Quantitative structure and aldose reductase inhibitory activity relationship of 1,2,3,4-tetrahydropyrrolo[1,2-*a*] pyrazine-4-spiro-3′-pyrrolidine-1,2′,3,5′- tetrone derivatives. Bioorg Med Chem 13:1445–1452
19. Bruno G, Costantino L, Curinga C et al (2002) Synthesis and aldose reductase inhibitory activity of 5-arylidene-2,4-thiazolidinediones. Bioorg Med Chem 10:1077–1084
20. Maccari R, Ottana R, Curinga C et al (2005) Structure-activity relationships and molecular modelling of 5-arylidene-2,4-thiazolidinediones active as aldose reductase inhibitors. Bioorg Med Chem 13:2809–2823
21. Sambasivarao SV, Soni LK, Gupta AK et al (2006) Quantitative structure-activity analysis of 5-arylidene-2,4-thiazolidinediones as aldose reductase inhibitors. Bioorg Med Chem Lett 16:512–520
22. Maccari R, Ottanà R, Ciurleo R et al (2007) Evaluation of in vitro aldose redutase inhibitory activity of 5-arylidene-2,4-thiazolidinediones. Bioorg Med Chem Lett 17:3886–3893
23. Rakowitz D, Maccari R, Ottanà R et al (2006) In vitro aldose reductase inhibitory activity of 5-benzyl-2,4-thiazolidinediones. Bioorg Med Chem 14:567–574
24. Pan HJ, Lin Y, Chen YE et al (2006) Adverse hepatic and cardiac responses to rosiglitazone in a new mouse model of type 2 diabetes: relation to dysregulated phosphatidylcholine metabolism. Vasc Pharmacol 45:65–71
25. ÓMoore-Sullivan TM, Prins JB et al (2002) Thiazolidinediones and type 2 diabetes: new drugs for an old disease. Med J Australia 176:381–386
26. Maccari R, Ottanà R, Curinga C, Vigorita MG et al (2005) Structure-activity relationships and molecular modelling of 5-arylidene-2,4-thiazolidinediones active as aldose reductase inhibitors. Bioorg Med Chem 13:2809–2823
27. Cramer RD, Patterson DE, Bunce JD (1988) Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967
28. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146
29. SYBYL Molecular Modeling Software, Version 6.9, Tripos Inc., St. Louis, MO
30. Benabadji HS, Chen HF et al (2003) 3D-QSAR study on diindolylmethane and its analogues with comparative molecular field analysis (CoMFA). Chinese J Chem 21:20–24
31. Wellsow J, Machullab HJ, Kovar KA (2002) 3D QSAR of serotonin transporter ligands: CoMFA and CoMSIA studies. Quant Struct-Act Relatsh 21:577–589
32. Golbraikh A, Tropsha A (2002) Beware of q^2! J Mol Graph Model 20:269–276
33. Golbraikh A, Tropsha A (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des 16:357–369
34. Golbraikh A, Tropsha A (2000) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. Mol Divers 5:231–243

35. Afantitis A, Melagraki G, Sarimveis H et al (2006) A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. Mol Divers 10:405–414
36. Roy PP, Roy K (2008) On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 27:302–313
37. Murthy VS, Kulkarni VM (2002) 3D-QSAR CoMFA and CoMSIA on protein tyrosine phosphatase 1B inhibitors. Bioorg Med Chem 10:2267–2282
38. Equbal T, Silakari O, Ravikumar M (2008) Exploring three-dimensional quantitative structural activity relationship (3D-QSAR) analysis of SCH 66336 (Sarasar) analogues of farnesyltransferase inhibitors. Eur J Med Chem 43:204–209
39. Subbiah Ramar SB, Tawari NR, Degani MS (2007) 3-D-QSAR analysis of 2-(oxalylamino) benzoic acid class of protein tyrosine phosphatase 1B inhibitors by CoMFA and Cerius2 GA. QSAR Comb Sci 26:608–617