



In silico studies of novel scaffold of thiazolidin-4-one derivatives as anti-*Toxoplasma gondii* agents by 2D/3D-QSAR, molecular docking, and molecular dynamics simulations

Rahman Abdizadeh¹ · Farzin Hadizadeh² · Tooba Abdizadeh³

Received: 29 May 2019 / Accepted: 12 November 2019 / Published online: 3 January 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Toxoplasma gondii is an obligate intracellular protozoa that can infect a wide variety of warm-blooded animals and humans. It was claimed that novel anti-*Toxoplasma gondii* agents were optimized as potential drug candidates, designed and created as significant agents. In this work, molecular modeling studies, including CoMFA, CoMFA-RF, CoMSIA, and HQSAR were performed on a set of 59 thiazolidin-4-one derivatives as anti-*T. gondii* agents. The statistical qualities of generating models were justified by internal and external validation, i.e., cross-validated correlation coefficient (q^2), non-cross-validated correlation coefficient (r_{ncv}^2) and predicted correlation coefficient (r_{pred}^2), respectively. The CoMFA (q^2 , 0.897; r_{ncv}^2 , 0.933; r_{pred}^2 , 0.938), CoMFA-RF (q^2 , 0.900; r_{ncv}^2 , 0.935; r_{pred}^2 , 0.998), CoMSIA (q^2 , 0.910; r_{ncv}^2 , 0.950; r_{pred}^2 , 0.998), and HQSAR models (q^2 , 0.924; r_{ncv}^2 , 0.953; r_{pred}^2 , 0.995) for training and test set yielded significant statistical results. Therefore, these QSAR models were excellent, robust, and had better predictive capability. Contour maps of the QSAR models were generated and validated by molecular dynamics simulation-assisted molecular docking study. The final QSAR models could be useful for the design and development of novel potent anti-*T. gondii* agents.

Keywords *Toxoplasma gondii* · 3D-QSAR · HQSAR · Thiazolidin-4-one derivatives · ROP8 · Molecular dynamics simulations

Highlights

- Multi-QSAR modeling studies were performed on thiazolidin-4-one derivatives as anti-*Toxoplasma gondii* agents.
- Molecular docking and molecular dynamics simulation were carried out to evaluate the interactions between the thiazolidin-4-one derivatives and ROP8 protein.
- The obtained results from QSAR and molecular docking studies were used to design novel potent anti-*Toxoplasma gondii* agents.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11224-019-01458-y>) contains supplementary material, which is available to authorized users.

✉ Tooba Abdizadeh
t.abdizadeh@gmail.com

- ¹ Department of Medical Parasitology and Mycology, Faculty of Medicine, Shahrekord University of Medical Sciences, Shahrekord, Iran
- ² Biotechnology Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran
- ³ Clinical Biochemistry Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran

Abbreviations

<i>T. gondii</i>	<i>Toxoplasma gondii</i>
ROPs	Rhoptry proteins
QSAR	Quantitative structure-activity relationship
3D-QSAR	Three-dimensional QSAR
CoMFA	Comparative molecular field analysis
CoMFA-RF	CoMFA region focusing
CoMSIA	Comparative molecular similarity index analysis
HQSAR	Hologram QSAR
MD	Molecular dynamics
AD	Application domain.

Introduction

Toxoplasma gondii is an obligate intracellular coccidian parasite of the phylum Apicomplexa that can infect any nucleated cell of a very wide range of warm-blooded vertebrates, including, carnivores, omnivores, and herbivores, as well as humans throughout the world, causing

toxoplasmosis [1, 2]. According to global investigation estimated, approximately 30% of the human population is chronically infected with toxoplasmosis specifically in developing countries. *T. gondii* in different hosts such as humans according to parasite strain and the immune status of the host cause infection with a range from asymptomatic to severe manifestations, therefore toxoplasmosis in immunocompetent persons is usually asymptomatic or associated with self-limited symptoms and rarely needs treatment [3, 4], but in immunocompromised individuals with primary or acquired deficiencies in T cell function such as AIDS patients and patients undergoing therapies for malignancies, transplants, or lymphoproliferative disorders because of reactivation of a latent infection, it may cause severe diseases ranging from encephalitis, myocarditis, pneumonia, hepatitis, and severe ocular disease to death [5]. Also in fetuses and children, congenital toxoplasmosis according to infected time of seronegative pregnant mothers may cause death of the fetus in the uterus or disorders such as ocular and neurologic complications in surviving infants [6]. Furthermore, *T. gondii* causes considerable economic loss and damage to the livestock industry, mainly among food-producing animals. Therefore, this protozoan has great importance in both public health and livestock husbandry. Nevertheless, so far, fully effective, safe, and efficacious vaccine is not produced for prevention of transmission and reducing serious complications of toxoplasmosis in humans and animals. The therapy for this disease has not changed in recent years, despite the tragic consequence of toxoplasmosis in humans. The current treatment for toxoplasmosis is a combination of antifolates (pyrimethamine and trimethoprim with a sulfonamide drug (sulfadiazine)) that display numerous and serious side effects such as hypersensitivity, hematological toxicity, teratogenicity, and allergic reactions [7] (Fig. 1).

So, the discovery and development of less-toxic and more-efficacious parasite-specific compounds becomes crucial for blocking any stage of the parasite's life cycle in humans or in a different type of hosts [8]. Among the heterocyclic systems, thiazolidine-4-one is a biologically important scaffold for potential drugs and drug candidates such as antiviral, antibacterial, antifungal, antihistamine,

and anti-inflammatory drugs. The thiazolidin-4-one analogs present promising pharmacological properties not only for the treatment of *T. gondii* infections but also for its high selectivity level with a high therapeutic index [9, 10].

The apical complex structure of *T. gondii* has served essential functions in both invasions of its host cells by attachment or penetration and in replication of the parasite. This structure includes three important secretory organelles known as rhoptries, micronemes, and dense granules. Rhoptries are morphologically club shaped with anterior duct (neck) and the posterior bulb that contain a conserved serine/threonine protein kinase domain and approximately constitute 1–30% of the total *Toxoplasma* cell volume. The rhoptry proteins (ROPs) have an important role in the multiple stages of the *T. gondii* invasion and also critical for survival within host cells (e.g., family ROP2, ROP4, ROP7, and ROP8) [11]. ROP8 is considered a type I transmembrane and has a conserved serine/threonine (S/T) kinase domain and contain in its cytoplasmic tails both tyrosine-based and dileucine-sorting signals. This protein expressed in the bradyzoites and tachyzoites stages of *T. gondii*, therefore, is a main factor of *T. gondii* acute virulence and has a key role in the parasitophorous vacuole (PV) formation [12].

Quantitative structure-activity relationship (QSAR) is a technique that is used in computer-assisted rational drug design and predicts the protein-ligand interaction and to explore the correlation between biological activity and molecular structure [13–15]. Three-dimensional QSAR (3D-QSAR) is a broad term encompassing all those QSAR methods which are utilized to calculate the highly specific interactions and a molecule, how far and with how much power can be connected to the active site of an enzyme or protein [16, 17]. Recently, comparative molecular field analysis (CoMFA), CoMFA region focusing (CoMFA-RF), comparative molecular similarity index analysis (CoMSIA), and hologram QSAR (HQSAR) are especially effective methods of QSAR based on statistical techniques [18–23].

In the present study, we performed a molecular modeling study by combining 2D- and 3D-QSAR, molecular docking, and molecular dynamics (MD) simulation techniques. 2D-QSAR, using HQSAR method, and 3D-QSAR, using CoMFA, CoMFA-RF, and CoMSIA methods, were used to identify the key structural factors influencing inhibitory activity. Molecular docking was used to identify some key amino acid residues at the active site of ROP8 protein and investigate the binding modes between ROP8 and the selected inhibitors. MD simulations were employed to determine the detailed interactions in ROP8 protein and validate the rationality of docking results. The obtained results can apply to the

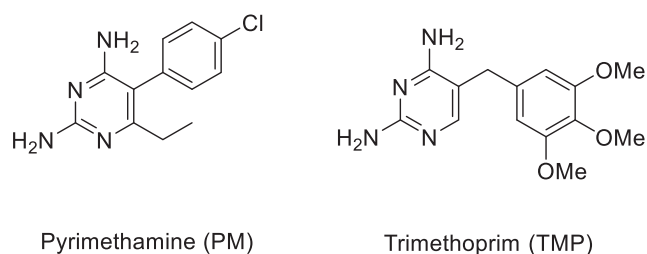
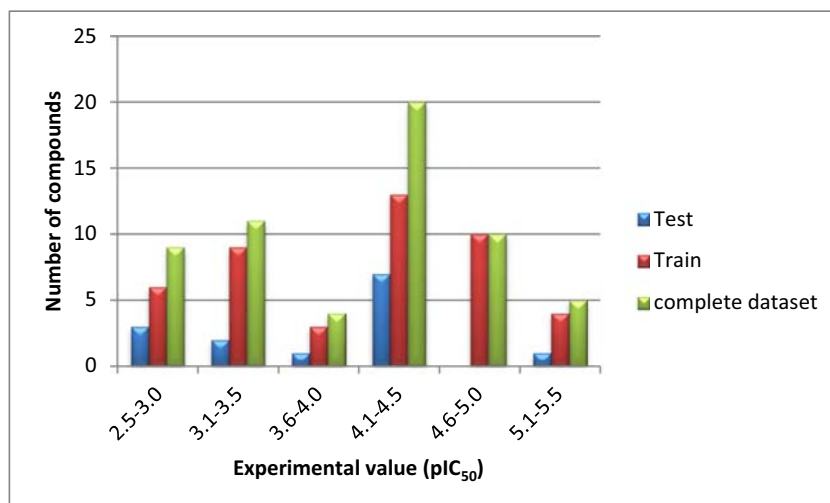


Fig. 1 DHFR inhibitors in the clinic for *T. gondii* infections

Fig. 2 Distribution of experimental inhibitory activities (pIC_{50}) for the training and test sets compounds in the QSAR models



further structural modification, design and develop new and more potent anti-toxoplasma drugs.

Materials and methods

Data set

QSAR studies were performed on a set of 59 thiazolidin-4-one derivatives as a new class of anti-*T. gondii* agents with their biological activities (IC_{50} values) that are recently reported by Tenorio, de Aquino, Carvalho, Liesen, and Carradori groups [9, 24–27].

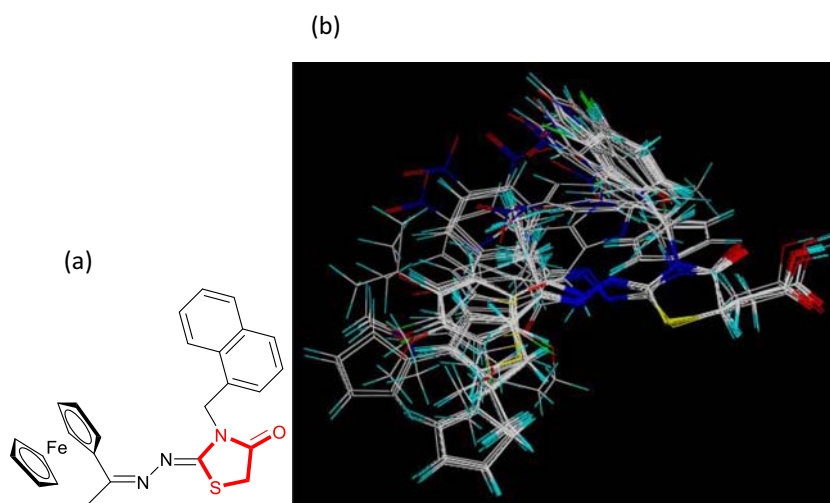
These activity values (IC_{50} in μM) were converted to corresponding pIC_{50} ($-\log IC_{50}$) values and used as a dependent variable in CoMFA, CoMFA-RF, CoMSIA, and HQSAR models. The data set was randomly divided into a training set (44 compounds, 75%) for QSAR model generation and a

test set (15 compounds, 25%) for external validation of the models (Fig. 2).

Molecular modeling and alignment

The QSAR models including CoMFA, CoMFA-RF, CoMSIA, and HQSAR were performed using the SYBYL-X 1.2. molecular modeling software (Tripos, Inc., St. Louis, MO). Before modeling with these primary methods, the 3D structures of compounds were drawn using Chemoffice Bio 3D Ultra (version 12.0, Cambridge Soft Corporation, Cambridge, UK, 2010). All the compounds were energy minimized using the standard molecular mechanics force field with a distance dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.05 kcal/molÅ using the maximum iteration set to 5000 [28]. Partial atomic charges of the compounds for electrostatic interactions were calculated by the Gasteiger-Hückel method. Structure alignment was one of the most important input variables in

Fig. 3 Compound **56** used as the template molecule for database alignment and common substructure in distill alignment shown in the bold red (a) and aligned compounds in the training and test sets (b)



3D-QSAR analysis, and the accuracy of the prediction power of the models was reliability dependent on contour maps according to the structural alignment of the molecules. In this study, rigid body alignment of molecules in a Mol2 database was performed using maximum common substructures defined by Distill alignment. Compound **56** was selected as template because the most active compound of the data set and other compounds were aligned according to the common structure. The structure of compound **56** with bold red common substructure and final super imposition of compounds are shown in Fig. 3a, b.

CoMFA and CoMSIA analysis

The CoMFA model (by Cramer et al.) describes the molecular properties by steric (Lennard-Jones) and electrostatic (Coulomb) energy fields of important regions of a set of aligned compounds that predict their biological activity over a lattice of point [29–31]. In CoMFA-RF model, steric and electrostatic fields are calculated for aligned fragments by creating specific grid space at the specific lattice points [32]. In CoMFA method, the aligned molecules in optimal orientation were located in a 3D cubic lattice with grid spacing of 2 Å in the *x*, *y*, and *z* direction which extended 4.0 Å around the align molecules in all Cartesian directions. The CoMFA steric and electrostatic fields were calculated for each molecule using a hybridized sp³ carbon probe atom with a Vander Waals radius of 1.52 Å and a charge of + 1.0. The Coulomb and Lennard-Jones potential functions were used to estimate the electrostatic and steric interactions, respectively. The energy cutoff values for both steric and electrostatic fields were set at 30 kcal/mol. In order to reduce noise and improve efficiency, column filtering was tested in the range of 0.0 to 2.0 kcal/mol and a threshold column filtering value of 2.0 kcal/mol. CoMFA-RF in the “Advanced CoMFA” module is a technique of application of weight to the lattice point in a CoMFA region to increase or decrease the contribution of these points to subsequent analysis. “StDev*Coefficients” values as different weighting factors were employed in addition to grid spacing for getting the better models. This increases the resolution and predictive capability (q^2 , cross-validated r^2) of a followed partial least squares (PLS) analysis [33].

In the CoMSIA model, proposed by Klebe et al., a probe atom is used to calculate similarity indices, at regularly placed grid points for the aligned molecules. Compared with CoMFA, CoMSIA uses a Gaussian-type distance-dependent function to assess five fields of different physicochemical properties (i.e., steric, electrostatic, hydrophobic, hydrogen-binding donor, and acceptor [34, 35]). Also, CoMSIA is differentiated by distance-related Gaussian functions and no arbitrary definitions of cut off limits should be used.

The CoMSIA method calculates the similarity indices descriptors with the same lattice box used in CoMFA. Five

physicochemical properties of steric, electrostatic, hydrophobic, hydrogen-binding donor, and acceptor fields were evaluated using a probe atom with to charge + 1.0, radius 1 Å, hydrophobicity + 1.0, hydrogen-binding donor + 1.0, hydrogen-binding acceptor + 1.0, attenuation factor α of 0.3 and grid spacing 2.0 Å. A distance-dependent Gaussian type was used between the probe atom and each molecule atom [28, 36]. In these models, all regression analyses performed in two steps using the PLS method [37–42].

HQSAR analysis

Hologram QSAR study is a 2D-QSAR technique which provides certainty to the relationship between the biological activity with the structural fragments which employs the fragment fingerprints of molecular holograms and other molecular descriptors to predict the biological activity of a series of molecules [43–45]. Hologram QSAR study is a 2D-QSAR technique which certain the relationship between the biological activity with the structural fragments. This method eliminates the need for 3D structure, the ability to achieve molecular alignment and conformational specification [46, 47] by transforming the chemical representation of a molecule into its corresponding molecular hologram. 2D chemical database storage and searching technologies rely on linear notations that define chemical structures (Wiswesser line-formula notation (WLN), simplified molecular input line entry system (SMILES); SLN-SYBYL line notation). The process involves the generation of fragments that are hashed into the array called molecular hologram, and bin occupancies are the descriptor variable [46, 48].

The HQSAR method employs different parameters for the molecular hologram generation, such as hologram length (HL) values (53, 59, 61, 72, 83, 97, 151, 199, 257, 307, 353, and 401), a fragment distinction (atom (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA), and the fragment size (2–5, 3–6, 4–7, 5–8, 6–9, 7–10).

Partial Least-Square analysis

In 3D-QSAR studies, PLS method [19, 49] which is an extension of multiple regression analysis was used for the model building. Calculated CoMFA and CoMSIA descriptors as independent variables were used with the pIC₅₀ values as dependent variables in the PLS regression analysis, respectively. Before the PLS analysis, the CoMFA and CoMSIA columns were filtered by using column-filtering value equal to 2.0 kcal/mol. The predictive ability of the models was evaluated by leave-one-out (LOO) and leave-ten-out (L-10-O) methods. LOO cross-validation method was used as an internal

validation to determine the number of components that yields an optimal predictive model. A component is a linear combination of the explanatory, independent column data. Unlike the independent explanatory variables themselves, the components are mutually orthogonal to one another. This number of components can only be as large as the number of degrees of freedom in the dataset. The optimal setting for un-cross-validated analyses is generally determined by preliminary cross-validated runs. The default setting of 6 is a reasonable place to start for cross-validation analyses.

The optimal number of principal components (ONC) was obtained by applying leave-one-out cross-validation, which was then utilized to derive the final QSAR models. The optimal number of components corresponds to one that produces the lowest standard error of prediction (SEP) and the highest cross-validated coefficient q^2 (r_{cv}^2) that was calculated by Eq. (1):

$$q^2 = 1 - \frac{\sum_{n=1}^{\infty} (\hat{y}_i - y_i)^2}{\sum_{n=1}^{\infty} (y_i - \bar{y})^2} \quad (1)$$

whereas, \hat{y}_i and y_i are predicted, observed activity values, and \bar{y} and \hat{y} are observed and predicted mean activity values of the training set, respectively [50]. The $\sum_{n=1}^{\infty} (\hat{y}_i - y_i)^2$ is the predictive residual sum of squares (PRESS).

After cross-validation, the final PLS analysis was carried out using the optimal number of components with no validation to generate the final QSAR model. The non-cross-validated analysis performed by the conventional correlation coefficient r^2 (r_{ncv}^2) (Eq. (2)), standard error of estimation (SEE) and F values calculated with the same column filtering set. High q^2 and r^2 ($q^2 > 0.5$, $r^2 > 0.6$) values are regarded as a proof of high predictive ability of the built model and also $r^2 - q^2$ for a good model should not be more than 0.3 [51].

$$r^2 = \frac{\left[\sum (y_i - \bar{y}_i) (\hat{y}_i - \hat{y}) \right]^2}{\sum (y_i - \bar{y}_i)^2 \times \sum (\hat{y}_i - \hat{y})^2} \quad (2)$$

Bootstrapping analysis was performed for 100 runs to assess the statistical confidence of the derived models [29, 49, 52, 53]. Contour maps were generated graphically after models were developed in CoMFA/CoMFA-RF and CoMSIA using the field type “StDev*Coeff”, and the contour levels were set to default values.

In HQSAR, LOO cross-validation was applied to determine the number of components that yields a good predictive model. PLS then yields a mathematical equation that related the molecular hologram bin values to the inhibitory activity of the compounds in the database.

Validation of the QSAR model

A good internal validation showed only a high q^2 in the training set of compounds, but it did not indicate the high predictive ability of the established models, therefore external validation was essential. The predictive ability of 3D-QSAR models was validated by calculating biological activities of the compounds which were not included in the training set and used as a test set. Test set was marked with an asterisk in Table 1.

The predictive correlation coefficient r_{pred}^2 ($r_{pred}^2 > 0.6$) [54], based on the test set, was calculated using Eq. (3):

$$r_{pred}^2 = \left(\frac{SD - PRESS}{SD} \right) \quad (3)$$

SD is the sum of squared deviation between the biological activities of the test set molecules and the mean activity of the training set molecules. PRESS is the sum of squared derivations between the predicted and actual activities of the test set molecules.

The performance of the regression models constructed here was evaluated using the root mean squared error (RMSE), mean absolute error (MAE) (RMSE and MAE close to zero), residual sum of squares (RSS), and concordance correlation coefficient (CCC; $CCC \geq 0.85$) of the training and validation sets [55]. The RMSE and the MAE are calculated for the data set as Eqs. (4)–(7):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (5)$$

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$CCC = \frac{2 \sum_{i=1}^n (y_i - \bar{y}) (\hat{y}_i - \hat{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - \hat{y})^2 + n (\bar{y} - \hat{y})^2} \quad (7)$$

To obtain the best predictive model for the test set, additional validation of model, the following parameters [54] were used (Eq. (8)):

$$\frac{(r^2 - r_o'^2)}{r^2} < 0.1 \text{ or } \frac{(r^2 - r_o'^2)}{r^2} < 0.1 \quad (8)$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15$$

r_0^2 and $r_o'^2$ are squared correlation coefficients of determination for regression lines through the origin between predicted (y) and observed (x) activities and vice versa. The values of k and k' are the slopes of their models, respectively.

Table 1 Chemical structure and the corresponding experimental and predicted pIC₅₀ values by QSAR models

Compd	R ₁ , R ₂ , R ₃	Experimental pIC ₅₀	Predicted pIC ₅₀			
			CoMFA	CoMFA-RF	CoMSIA	HQSAR
1		3.699	3.455	3.628	3.433	3.526
2		3.301	3.445	3.623	3.426	3.447
3*		3.155	3.328	3.210	3.016	3.199
4		3.301	3.420	3.597	3.422	3.389
5*		3.301	3.124	3.249	3.571	3.328
6		3.341	3.527	3.676	3.521	3.364
5*		3.301	3.124	3.249	3.571	3.328
6		3.341	3.527	3.676	3.521	3.364

Table 1 (continued)

7*		3.00	2.906	3.127	3.153	3.438
8		3.301	3.154	3.094	3.113	3.080
9		3.00	3.123	3.083	3.079	3.036
10*		2.699	3.070	3.111	2.817	2.927
11		3.311	3.171	3.113	3.145	3.079
12		3.00	3.180	3.126	3.120	3.191
13*		4.301	4.182	4.112	4.213	4.593

Table 1 (continued)

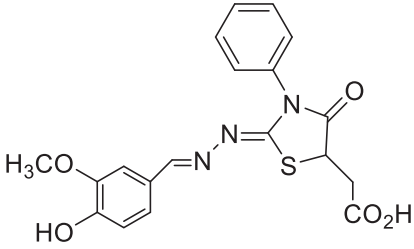
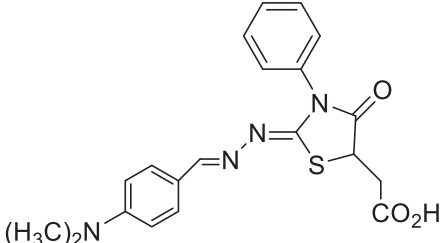
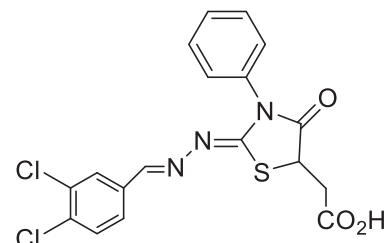
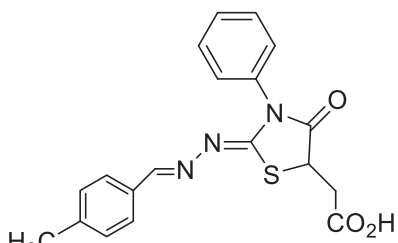
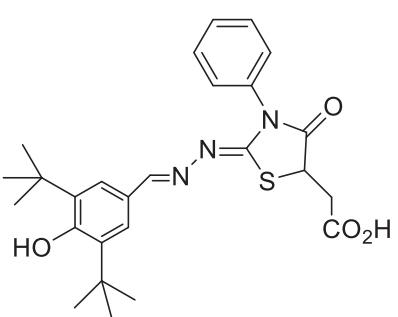
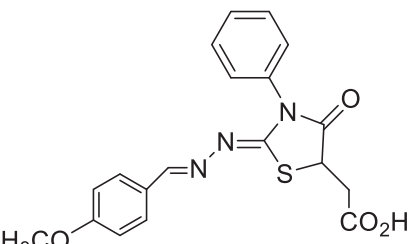
14		3.00	3.100	3.068	3.053	3.226
15*		4.301	4.123	4.091	4.265	4.501
16		3.00	3.128	3.077	3.074	3.037
17*		4.301	4.171	4.111	4.247	4.204
18		3.00	3.138	3.141	3.103	2.956
19		3.301	3.137	3.089	3.141	3.210

Table 1 (continued)

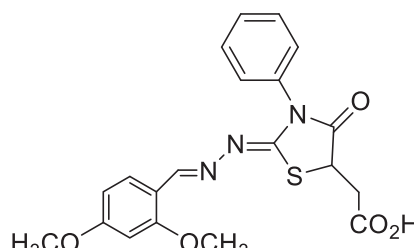
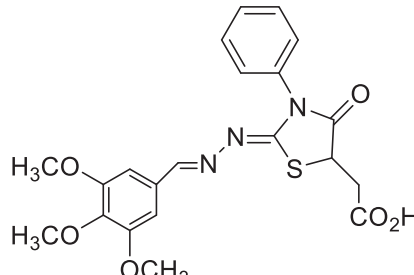
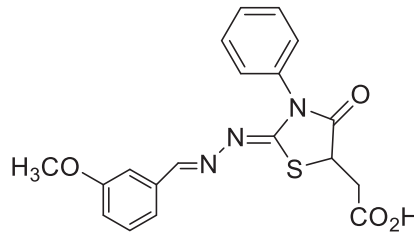
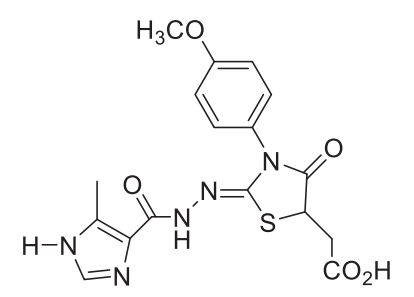
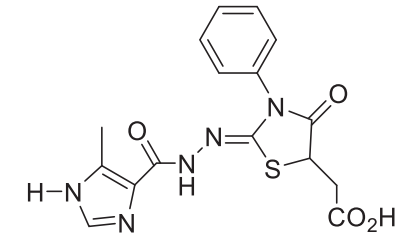
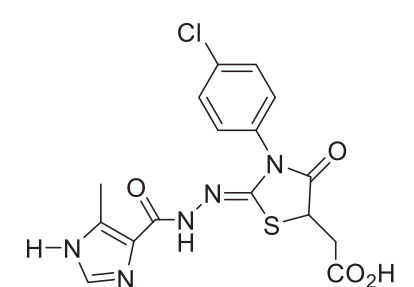
20		3.341	3.145	3.108	3.147	3.254
21		3.301	3.167	3.125	3.120	3.294
22*		4.00	4.136	4.097	4.233	4.225
23		3.371	3.070	2.991	3.182	3.210
24*		2.894	3.085	3.124	3.063	3.024
25		2.824	3.051	2.979	3.134	3.024

Table 1 (continued)

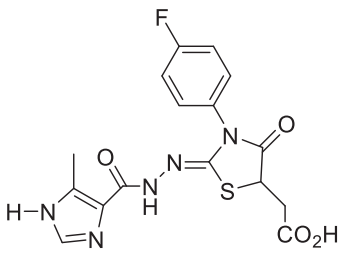
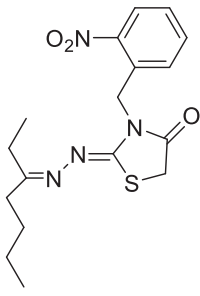
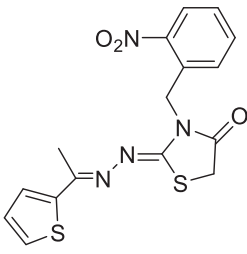
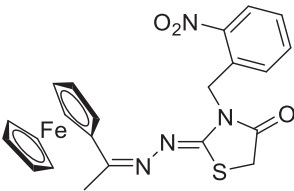
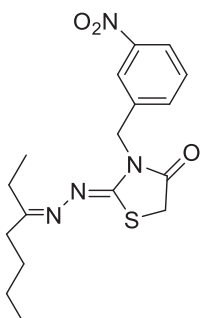
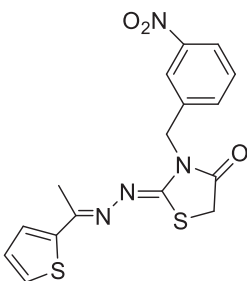
26*		4.00	4.062	3.983	4.163	4.398
27		4.699	4.479	4.539	4.543	4.558
28*		4.432	4.119	4.269	4.320	4.304
29		4.824	5.034	4.953	4.916	5.010
30		4.357	4.50	4.574	4.525	4.409
31		4.284	4.092	4.10	4.184	4.149

Table 1 (continued)

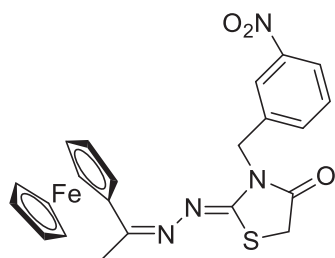
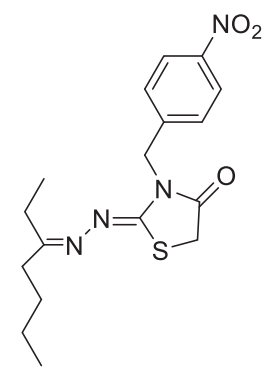
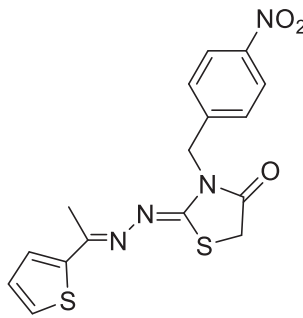
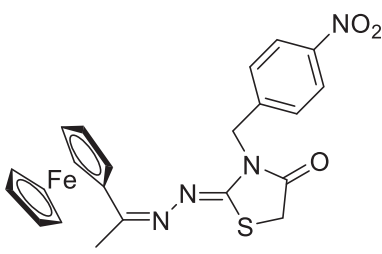
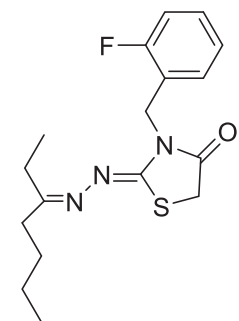
32		4.620	5.001	4.917	4.728	4.813
33		4.495	4.457	4.531	4.524	4.373
34		3.830	4.136	4.163	4.20	4.113
35*		4.137	4.004	4.228	4.189	4.207
36		4.620	4.487	4.562	4.544	4.581

Table 1 (continued)

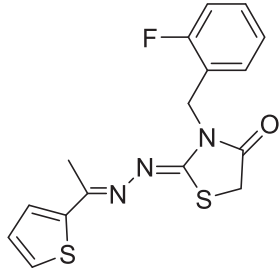
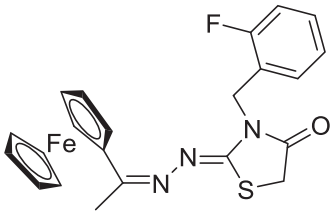
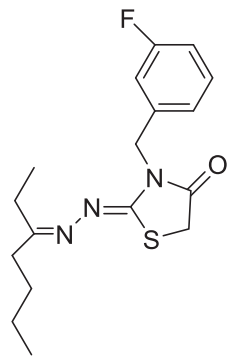
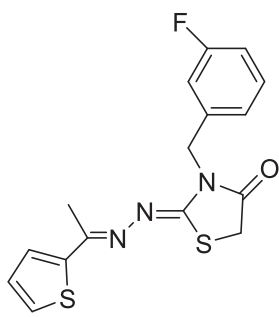
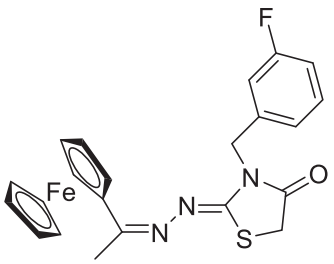
37		4.602	4.333	4.229	4.289	4.321
38		4.770	5.004	4.921	4.993	5.018
39		4.745	4.481	4.559	4.556	4.595
40		4.347	4.319	4.296	4.298	4.335
41		5.097	5.014	4.925	5.001	5.032

Table 1 (continued)

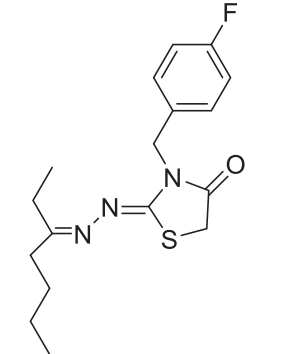
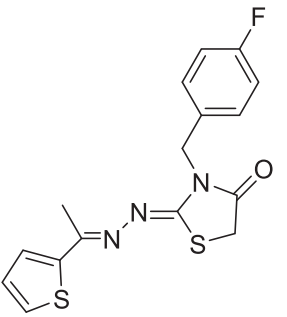
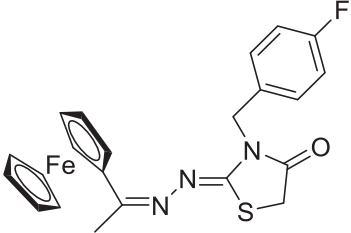
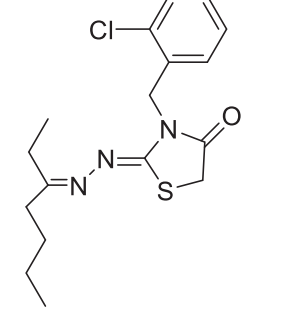
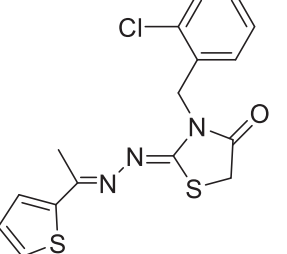
42		4.538	4.460	4.552	4.528	4.587
43		4.143	4.296	4.285	4.270	4.327
44*		5.22	5.112	5.122	5.10	5.014
45		4.409	4.456	4.521	4.542	4.489
46		4.469	4.308	4.275	4.288	4.228

Table 1 (continued)

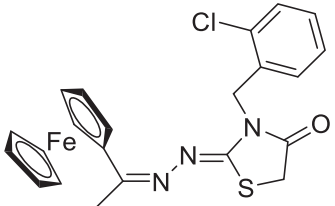
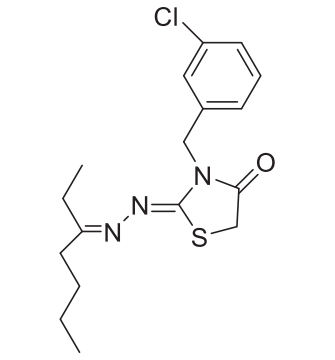
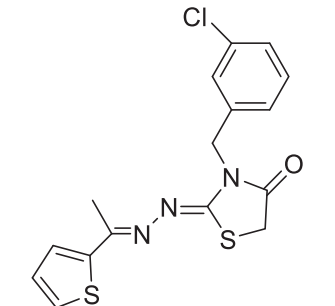
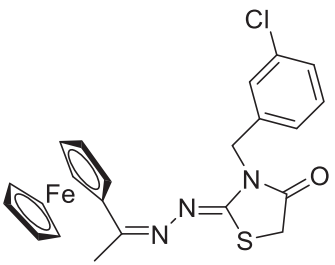
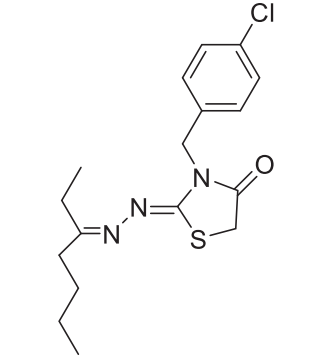
47		5.046	5.033	4.946	5.003	4.925
48		4.721	4.483	4.562	4.554	4.538
49		4.292	4.314	4.295	4.298	4.278
50		5.097	5.029	4.929	5.027	4.975
51		4.292	4.474	4.561	4.522	4.484

Table 1 (continued)

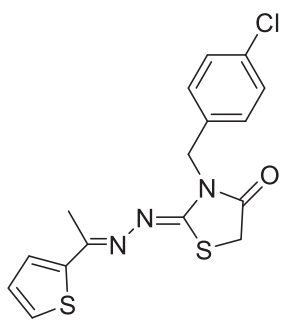
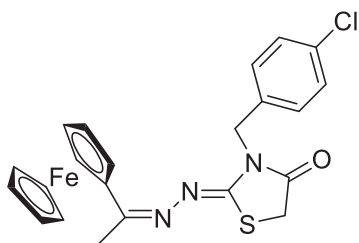
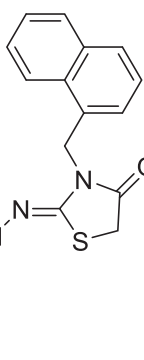
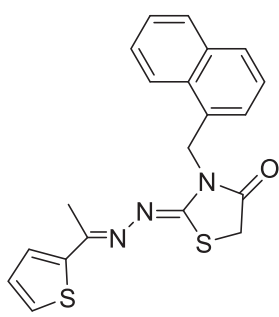
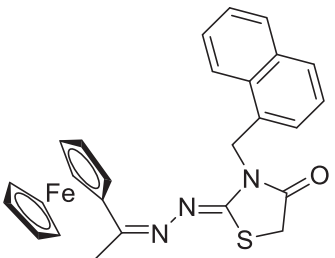
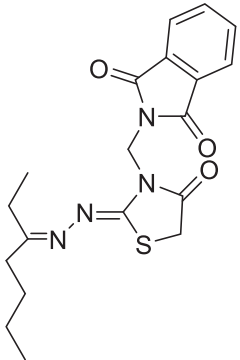
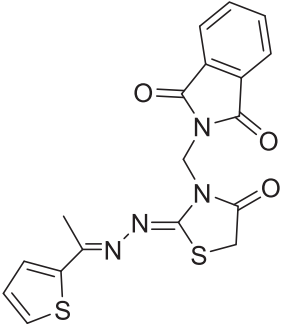
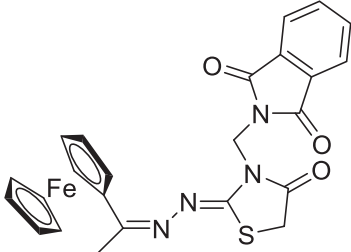
52		4.092	4.229	4.288	4.266	4.224
53		5.00	5.023	4.925	5.025	4.92
54		4.443	4.461	4.555	4.535	4.717
55		4.456	4.288	4.273	4.276	4.456
56		5.301	4.946	4.819	5.087	5.153

Table 1 (continued)

57*		4.056	4.216	4.108	4.125	3.964
58		4.108	4.306	4.285	4.191	3.990
59*		4.086	4.306	4.228	4.127	4.192

*Test set

To further assess the models, another validation statistical parameter r_m^2 and Δr_m^2 were determined by Eqs. (9) and (10):

$$r_m^2 = r^2 \left(1 - \sqrt{|r^2 - r_o^2|} \right) \quad (9)$$

$$\Delta r_m^2 = |r_m^2 - r'_m{}^2| \quad (10)$$

r_m^2 value more than 0.5 ($r_m^2 > 0.5$) and $\Delta r_m^2 < 0.2$ show good external predictability of the models.

Molecular docking study

Molecular docking as one of the most frequent methods in drug design was used to investigate the mode of interaction of small

molecules with the appropriate target binding sites. The docking study was performed using Operation Environment (MOE) software (www.chemcomp.com) between the most and least active compounds with TgROP8 protein. For the preparation of ligands prior to docking, the 2D structures of ligands were prepared by Chemoffice ultra (version 12.0, Cambridge Soft Corporation, Cambridge, UK, 2010) and converted to 3D format by Hyper Chem7 (Hyper cube Inc., USA) using AM1 semi-empirical method. The ligands in our data set were docked in the active site of TgROP8 (PDB ID: 3byv) by MOE software. The docking was performed by the triangle matcher placement algorithm in combination with London dG scoring function and force field as refinement method, and the conformation of compounds were further analyzed by LigX module in MOE software. The best docking pose of compound **56** was chosen for MD simulation.

Molecular dynamics simulation

The MD simulations, based on Newton's second law or the equation of motion, were performed to investigate the interaction between the receptor and ligand in atomic details using the dynamics module of SYBYL [56]. The compound **56** was used as the template molecule to elucidate the MD simulations. Energy minimization of the docked ligand was performed in the Tripos force field and Gasteiger-Huckel charge without water using Boltzmann initial velocity.

The simulations were executed using normal temperature and volume (NTV) ensemble 300 K with coupling 100 fs. The MD simulations were performed with a time step of 2 for 10,000 fs and conformation snapshot at every 100 fs to calculate RMSD, Rg and potential energy values and recorded as a plot.

Results and discussion

CoMFA and CoMFA-RF statistical results

The statistical results of CoMFA and CoMFA-RF models are summarized in Table 2. The CoMFA analysis was carried out with steric and electrostatic fields at column filtering of 2.0 kcal/mol.

PLS analysis of CoMFA for the training set including leave-one-out (LOO) and leave-ten-out (L-10-O) cross-

validation with ONC 3 showed q^2 value of 0.897, r_{cv}^2 (L-10-O) value of 0.892, and SEP of 0.242. These statistical results showed that the model had a better predictive capability.

The non-cross-validated PLS analysis gave a r_{ncv}^2 of 0.933 with SEE of 0.191, F value of 194.977, $r^2 - q^2$ of 0.036, and $R_{pearson}$ of 0.932 which supported the statistical validity of the development model. The contributions from steric and electrostatic field descriptors explained 0.709 and 0.291 of the total variance, respectively, that indicated steric effect was more important than the electrostatic fraction.

After using region focusing, a new model of CoMFA-RF with improvement in the statistical parameters was created. The cross-validation and non-cross-validated PLS calculation results were found better in CoMFA-RF as, compared with CoMFA. This approach showed an increase in the q^2 value from 0.897 to 0.900 with ONC of 2 and r_{cv}^2 (L-10-O) from 0.892 to 0.902 and SEP of 0.236. The non-cross-validated PLS analysis resulted in high r_{ncv}^2 value of 0.935 with a low SEE value of 0.174, F value of 242.439, $r^2 - q^2$ value of 0.035, and $R_{pearson}$ value of 0.940. The contribution of steric and electrostatic field descriptors was 0.730 and 0.270, respectively in CoMFA-RF.

The bootstrapped results were shown in r_{bs}^2 and SEE_{bs} values of 0.984 and 0.005 (CoMFA) and 0.980 and 0.006 (CoMFA-RF), respectively, suggesting a good internal consistency and the absence of systematic errors of the models within the training data set.

Table 2 Statistical parameters of QSAR models

Parameters	CoMFA	CoMFA-RF	CoMSIA	HQSAR
PLS analysis				
q^2	0.897	0.900	0.910	0.924
r_{cv}^2 (L-10-O)	0.892	0.902	0.912	0.896
ONC	3	2	3	4
SEP	0.242	0.236	0.226	0.210
r_{ncv}^2	0.933	0.935	0.950	0.953
SEE	0.191	0.174	0.169	0.162
$R_{pearson}$	0.932	0.940	0.951	0.954
F	194.977	242.439	252.617	226.360
r_{bs}^2	0.984	0.980	0.981	0.982
SEE_{bs}	0.005	0.006	0.006	0.006
Contribution				
Steric	0.709	0.730	0.143	–
Electrostatic	0.291	0.270	0.170	–
Hydrophobic	–	–	0.206	–
Donor	–	–	0.356	–
Acceptor	–	–	0.126	–
Steric (VIF)	1.287	1.477	1.746	–
Electrostatic (VIF)	1.287	1.477	2.971	–
Hydrophobic (VIF)	–	–	3.483	–
Donor (VIF)	–	–	1.465	–
Acceptor (VIF)	–	–	2.374	–

q^2 , cross-validated correlation coefficient after the leave-one-out procedure; ONC, optimal number of principal components; r_{ncv}^2 , non-cross-validated correlation coefficient; SEE: standard error of estimate; F , the value of F statistic; r_{bs}^2 , the average r^2 value from a bootstrapping analysis for 100 runs; SEE_{bs} , the average SEE value from a bootstrapping analysis for 100 runs; r_{cv}^2 , (mean) the average r_{cv}^2 , from ten times tenfold cross-validation

CoMSIA statistical results

The CoMSIA technique deals with direct correlation of ligand affinities to changes in molecular properties [30]. The CoMSIA model was generated using combinations of five steric (*S*), electrostatic (*E*), hydrophobic (*H*), hydrogen-binding acceptor (*A*), and hydrogen-bonding donor (*D*) fields. The statistical parameters of CoMSIA model are summarized in Table 2. In PLS analysis, the q^2 value of 0.910 with ONC of 3, SEP of 0.226, and r_{cv}^2 (L-10-O) of 0.912 was obtained with column filtering of 2.0 kcal/mol. The non-cross-validated PLS analysis gave a r_{ncv}^2 value of 0.950 with SEE value of 0.169, F value of 252.617, $r^2 - q^2$ value of 0.040, and $R_{pearson}$ value of 0.951.

A high bootstrapped r^2 value of 0.981 and SEE_{bs} of 0.006 suggests a high degree of confidence in the analysis. For CoMSIA, the contribution of the steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor field descriptors were 0.143, 0.170, 0.206, 0.356, and 0.126, respectively. These molecular fields were not completely independent of each other and could form 31 combinations (Fig. 4).

Among the first five models, acceptor field with a high q^2 value ($q^2 = 0.910$) was more important than the other four fields. In CoMSIA model, combination of steric, hydrophobic, and hydrogen bond acceptor (SHA) was found to be the best. CoMSIA (SHA) combination gave q^2 value of 0.932, r_{ncv}^2 of 0.951, r_{cv}^2 of 0.923, and r_{bs}^2 of 0.982. In the model CoMSIA, this combination shared the large part and indicated that internal prediction of SHA combination was good.

According to the results of the CoMFA and CoMSIA models, the steric field and the steric, hydrophobic and hydrogen bond acceptor contributions, respectively shared the large part. Also, docking study show that the steric, hydrophobic and H-bond effects around the key residues of the active site

performed a significant role in the binding of ligand to TgROP8. It was indicated that the hydrophobic and steric properties were effective in the design of new *T. gondii* agents.

HQSAR statistical results

The HQSAR is a technique for QSAR analysis that is useful in exploring the combination of each molecule under study to the biological activity and eliminates the need of alignment, generation of 3D structures and putative binding conformation. The performance of the HQSAR model was affected by three parameters, including the fragment size, the fragment type (fragment distinction), and hologram length. The HQSAR models with statistical parameters are shown in Table 2.

The best statistical results of HQSAR model were obtained with q^2 value of 0.924, ONC of 4, SEP of 0.210, and r_{cv}^2 (L-10-O) of 0.896, r_{ncv}^2 of 0.953 with SEE of 0.162, F value of 226.360, $r^2 - q^2$ of 0.029, r_{bs}^2 of 0.982 with SEE_{bs} of 0.006, and $R_{pearson}$ of 0.954 using a relevant hologram length (HL) of 151, fragment distinction (atom (A) and bond (B)), and the fragment size of 4–7 (Tables 3 and 4). All the results demonstrated that the HQSAR model was also highly predictive.

Validation of QSAR models

The predictive abilities of the QSAR models were externally validated using the independent test set that was not used for the model generation [50]. q^2 and r^2 parameters, obtained from internal validation, were used for confirming the stability and the predictive ability of the models. The QSAR models for the whole test set including 15 compounds gave the r_{pred}^2 and r_m^2 values of 0.938 and 0.771 (CoMFA), 0.988 and 0.725 (CoMFA-RF), 0.998 and 0.870 (CoMSIA), and 0.995 and 0.763 (HQSAR) and high slope regression lines with k and k' values of 0.995

Fig. 4 Results of the distribution of q^2 , r_{ncv}^2 , r_{cv}^2 , and r_{bs}^2 values that were obtained from 31 combinations of CoMSIA fields. *s*, steric; *e*, electrostatic; *h*, hydrophobic; *d*, H-bond donor; *a*, H-bond acceptor

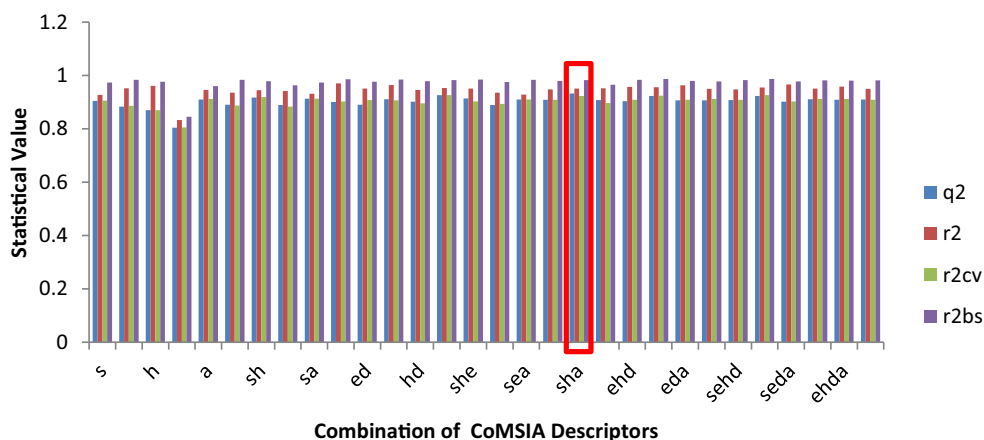


Table 3 HQSAR analysis for various fragment distinctions on the key statistical parameters using fragment size (4–7)

Model	Fragment distinction	q^2	SEP	r^2	SEE	HL	N
1–1	A	0.894	0.249	0.944	0.180	53	4
1–2	A/B	<i>0.924</i>	<i>0.210</i>	<i>0.953</i>	<i>0.162</i>	<i>151</i>	<i>4</i>
1–3	A/C	0.909	0.231	0.957	0.159	71	4
1–4	A/H	0.917	0.223	0.955	0.164	401	5
1–5	A/Ch	0.912	0.226	0.954	0.164	401	4
1–6	A/DA	0.825	0.328	0.944	0.185	53	6
1–7	A/B/C	0.907	0.238	0.959	0.159	59	6
1–8	A/B/H	0.910	0.230	0.944	0.181	307	4
1–9	A/B/Ch	0.922	0.211	0.953	0.164	307	3
1–10	A/B/DA	0.884	0.263	0.953	0.168	71	5
1–11	A/C/H	0.913	0.226	0.947	0.175	151	4
1–12	A/C/Ch	0.904	0.237	0.948	0.175	97	4
1–13	A/C/DA	0.873	0.275	0.954	0.166	61	5
1–14	A/H/Ch	0.918	0.225	0.960	0.156	257	46
1–15	A/H/DA	0.882	0.262	0.935	0.195	71	4
1–16	A/Ch/DA	0.845	0.304	0.943	0.185	151	5
1–17	A/B/C/H	0.913	0.231	0.961	0.155	307	6
1–18	A/B/C/Ch	0.901	0.240	0.956	0.159	199	4
1–19	A/B/C/DA	0.886	0.262	0.955	0.163	97	5
1–20	A/B/H/Ch	0.914	0.226	0.954	0.166	307	5
1–21	A/B/H/DA	0.905	0.235	0.952	0.167	151	4
1–22	A/B/Ch/DA	0.901	0.247	0.957	0.163	71	6
1–23	A/C/H/Ch	0.922	0.217	0.958	0.158	151	5
1–24	A/H/Ch/DA	0.872	0.273	0.939	0.188	97	4
1–25	A/C/H/DA	0.900	0.241	0.948	0.174	151	4
1–26	A/C/H/Ch/DA	0.908	0.231	0.950	0.171	199	4
1–27	A/B/H/Ch/DA	0.911	0.231	0.957	0.161	97	5
1–28	A/B/C/Ch/DA	0.907	0.233	0.951	0.168	97	4
1–29	A/B/C/H/DA	0.911	0.231	0.957	0.161	97	5
1–30	A/B/C/H/Ch	0.914	0.229	0.961	0.155	307	6
1–31	A/B/C/H/Ch/DA	0.902	0.239	0.950	0.170	97	4

The model chosen for analysis is set in italics

q^2 , cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; *SEE*, standard estimated error; *HL*, hologram length; *N*, optimal number of components. Fragment distinction: *A*, atom; *B*, bond; *C*, connections; *H*, hydrogen atom; *Ch*, chirality; *D*, donor and acceptor

and 1.002 (CoMFA), 1.000 and 0.997 (CoMFA-RF), 1.009 and 0.989, and 1.018 and 0.991 (HQSAR), respectively. r_o^2 and $r_o'^2$ values of 0.909 and 0.931 (CoMFA), 0.895 and 0.925 (CoMFA-RF), 0.952 and 0.961 (CoMSIA), and 0.897 and 0.839 (HQSAR), respectively, were used to calculate the relationship between r^2 , r_o^2 , and $r_o'^2$ that $(r^2 - r_o^2)/r^2$ and $(r^2 - r_o'^2)/r^2$ values of 0.035 and 0.012 (CoMFA), 0.060 and 0.025 (CoMFA-RF), 0.016 and 0.006 (CoMSIA), and 0.034 and 0.001 (HQSAR), respectively were obtained.

Table 4 Statistical results of QSAR model using the model 1–2 (including fragments A/B) with different fragment sizes

Model	Fragment size	q^2	SEP	r^2	SEE	HL	N
2–1	1–4	0.899	0.246	0.946	0.180	53	5
2–2	2–5	0.900	0.245	0.949	0.174	59	5
2–3	3–6	0.914	0.221	0.948	0.172	353	3
2–4	4–7	<i>0.924</i>	<i>0.210</i>	<i>0.953</i>	<i>0.162</i>	<i>151</i>	<i>4</i>
2–5	5–8	0.917	0.220	0.951	0.169	353	4
2–6	6–9	0.900	0.241	0.943	0.182	307	4
2–7	7–10	0.900	0.241	0.950	0.171	353	4
2–8	8–11	0.901	0.243	0.957	0.160	53	5
2–9	9–12	0.890	0.257	0.958	0.159	151	5

The model chosen for analysis is set in italics

q^2 , cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; *SEE*, standard estimated error; *HL*, hologram length; *N*, optimal number of components. Fragment distinction: *A*, atom; *B*, bond; *C*, connections; *H*, hydrogen atom; *Ch*, chirality; *D*, donor and acceptor

Table 5 Statistical parameters of validation method for QSAR models

Parameters	CoMFA	CoMFA-RF	CoMSIA	HQSAR
r_{pred}^2	0.938	0.998	0.998	0.995
r_o^2	0.909	0.895	0.952	0.897
$r_o'^2$	0.931	0.928	0.961	0.839
$(r^2 - r_o^2)/r^2$	0.035	0.060	0.016	0.034
$(r^2 - r_o'^2)/r^2$	0.012	0.025	0.006	0.001
k	0.995	1.000	1.009	1.018
k'	1.002	0.997	0.989	0.991
r_m^2	0.771	0.725	0.870	0.763
$r_m'^2$	0.843	0.805	0.892	0.813
Δr_m^2	0.072	0.080	0.022	0.05
RMSE _{train}	0.157	0.177	0.140	0.133
RMSE _{test}	0.089	0.091	0.157	0.136
MAE _{train}	0.119	0.132	0.111	0.098
MAE _{test}	0.040	0.038	0.032	0.056
RSS _{train}	1.454	1.859	1.161	1.046
RSS _{test}	0.465	0.487	1.455	1.100
CCC _{train}	0.967	0.888	0.951	0.976
CCC _{test}	0.961	0.955	0.968	0.904

r_{pred}^2 , predicted correlation coefficient for the test set of compounds; r^2 : correlation coefficient for regression through the origin for predicted vs. observed activities (Test set); $r_o'^2$: correlation coefficient for regression through origin for observed versus predicted activities (Test set); (r_m^2): modified squared correlation coefficient (test set); RMSE: root mean squared error; MAE: mean absolute error; RSS: residual sum of squares; CCC: concordance correlation coefficient.

The QSAR models yielded RMSE, MAE and CCC values of 0.157, 0.119, 0.967; 0.089, 0.040, and 0.961 (CoMFA); 0.177, 0.132, 0.888; 0.091, 0.038, and 0.955 (CoMFA-RF); 0.140, 0.111, and 0.951; 0.157, 0.032, and 0.968 (CoMSIA); and 0.133, 0.098, 0.976; 0.136, 0.056, and 0.904 (HQSAR) for training and test set, respectively.

From the values of the performance criteria parameters yielded by the QSAR models in training and test data (Table 5), it is evident that all of the models yielded considerably low RMSE and MAE values and high CCC values which show that models built by training set could be used for the prediction of these chemo types.

The results of external validation parameters are listed in Table 5. These results confirm that the QSAR models could be used to predict the biological activities of new compounds and their derivatives.

The correlation plots between the predicted and experimental activities. Most of the compounds were located on or near to the trend line in the QSAR models, and these results confirm that these models had good predictive ability for new compounds.

The residual values of the QSAR models. The CoMSIA and HQSAR models showed smaller residuals than the CoMFA and CoMFA-RF models and were the better models are shown in Fig. 5 and Fig. 6.

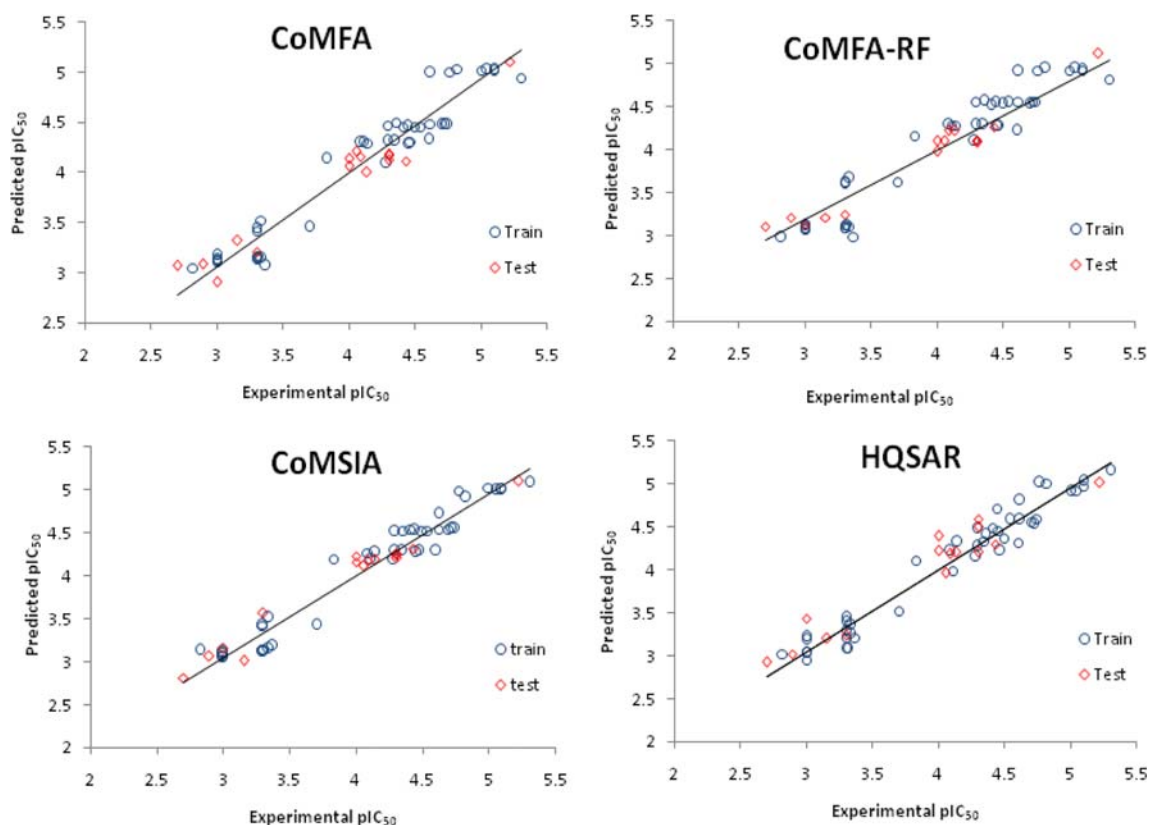


Fig. 5 The plot of predicted pIC_{50} vs. experimental pIC_{50} values for training and test sets compounds by QSAR models

Evaluation of the Y-randomization test, variance inflation factor, and application domain of model

The QSAR models were further validated by applying the Y-randomization test to assess the robustness of the models and to avoid chance correlation [57, 58]. Thus, for every original model, several random shuffles of the dependent variable (biological activity) were performed and a new QSAR model was developed using the original independent variable matrix and the results are shown in Table 6. The low q^2 and r_{ncv}^2 values ($q^2 < 0.5$ and $r_{ncv}^2 < 0.6$) show that the good results obtained in the formulation of the final models were not by chance.

In addition to Y-randomization tests, multi-collinearity of the descriptors and the models were detected by calculating the variance inflation factor (VIF), which can be calculated as follows:

$$VIF = \frac{1}{1-r^2}$$

where r^2 is the correlation coefficient of the multiple regression between the variables within the model. If VIF equals to 1, then no inter-correlation exists for each variable; if VIF falls into the range of 1–5, the related model is acceptable; and if VIF is larger than 10, the related model is unstable and a

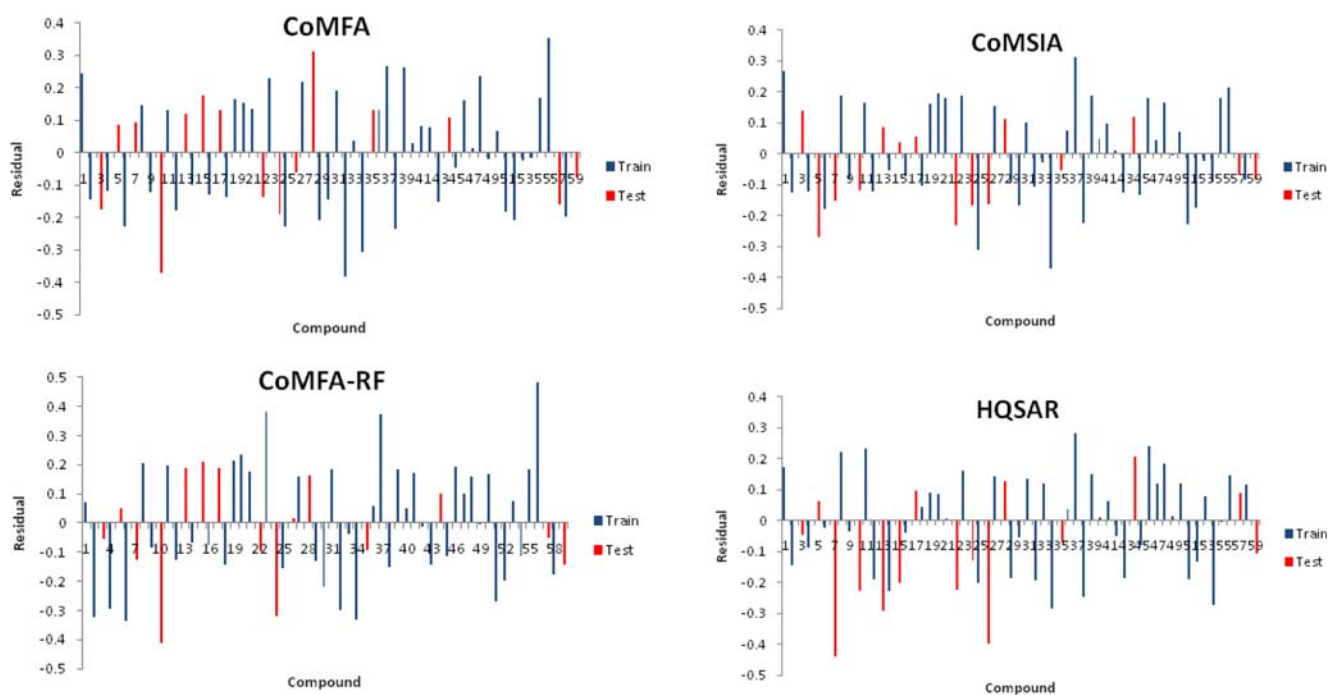


Fig. 6 Residual plots between experimental and predicted values for QSAR models

recheck is necessary [59, 60]. The corresponding VIF values of the seven descriptors were showed in Table 2. As can be

Table 6 q^2 and r_{ncv}^2 values after several Y-randomization tests

Y-random iteration	CoMFA		CoMSIA	
	q^2	r_{ncv}^2	q^2	r_{ncv}^2
1	0.304	0.460	0.246	0.378
2	0.316	0.436	0.242	0.364
3	0.282	0.345	0.256	0.275
4	0.273	0.387	0.236	0.343
5	0.181	0.377	0.275	0.381
6	0.210	0.411	0.241	0.263
7	0.248	0.421	0.315	0.380
8	0.231	0.395	0.246	0.255
9	0.295	0.356	0.263	0.324
10	0.208	0.376	0.197	0.269
11	0.248	0.368	0.201	0.365
12	0.351	0.379	0.258	0.278
13	0.180	0.424	0.295	0.308
14	0.278	0.405	0.279	0.313
15	0.292	0.443	0.281	0.351
16	0.265	0.387	0.237	0.378
17	0.270	0.358	0.265	0.367
18	0.240	0.369	0.258	0.363
19	0.261	0.367	0.272	0.376
20	0.233	0.359	0.205	0.351
Non-random	0.897	0.9331	0.910	0.950

seen from this table, all the variables have VIF values of less than five, indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal and model is said to be stably acceptable.

For a new compound with no experimental data, a predicted value of QSAR models without an idea of reliability of the value is not useful. Therefore, for evaluating new compounds, a very important step in QSAR model development is the definition of the applicability domain of regression or classification models [33].

The Williams plot, the plot of the standardized residuals (δ) vs. leverage values (h_i), was used to illustrate the predictive and express the applicability domain of the models for each chemical compound [61, 62].

The standardized residuals (δ) value is calculated by Eq. (11) [63]:

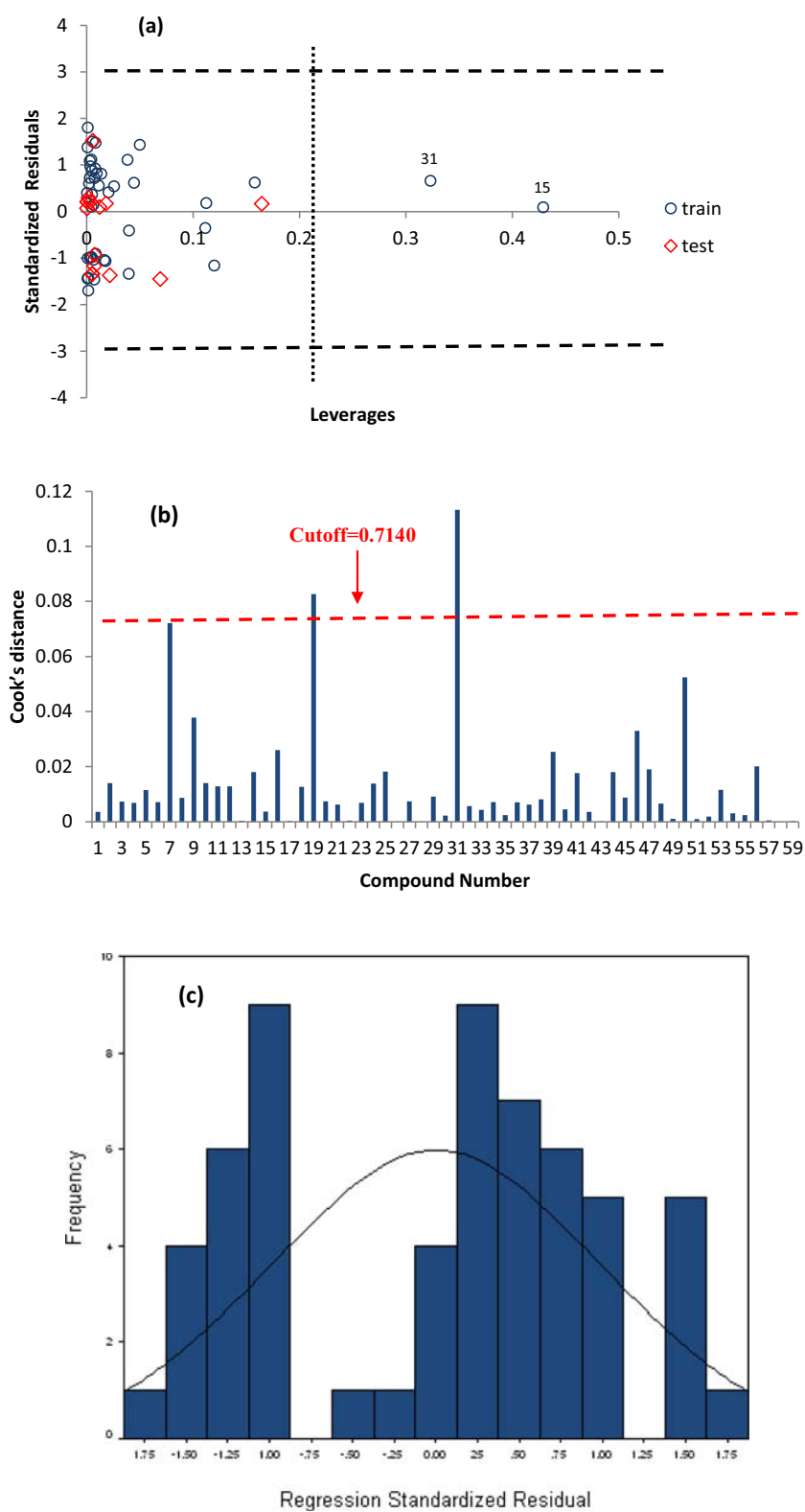
$$\delta = \frac{y_i - \hat{y}_i}{\sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{(n-A-1)}}} \quad (11)$$

where y_i and \hat{y}_i are the observed and predicted values for i th compound, respectively, n is the number of compounds, and A is the number of descriptors. Also, the leverage value (h_i) is defined by Eq. (12):

$$h_i = X_i^T (X^T X)^{-1} x_i \quad (i = 1, \dots, n) \quad (12)$$

where x_i is the descriptor-row vector of the i th compound, X_i^T is the transpose of x_i , X is the descriptor matrix of the training

Fig. 7 Williams plot describing the applicability domain of the CoMFA model for the training and test sets ($h^* = 0.206$) (a); Cook's distance plot (b); and histogram of model CoMFA residuals (c)



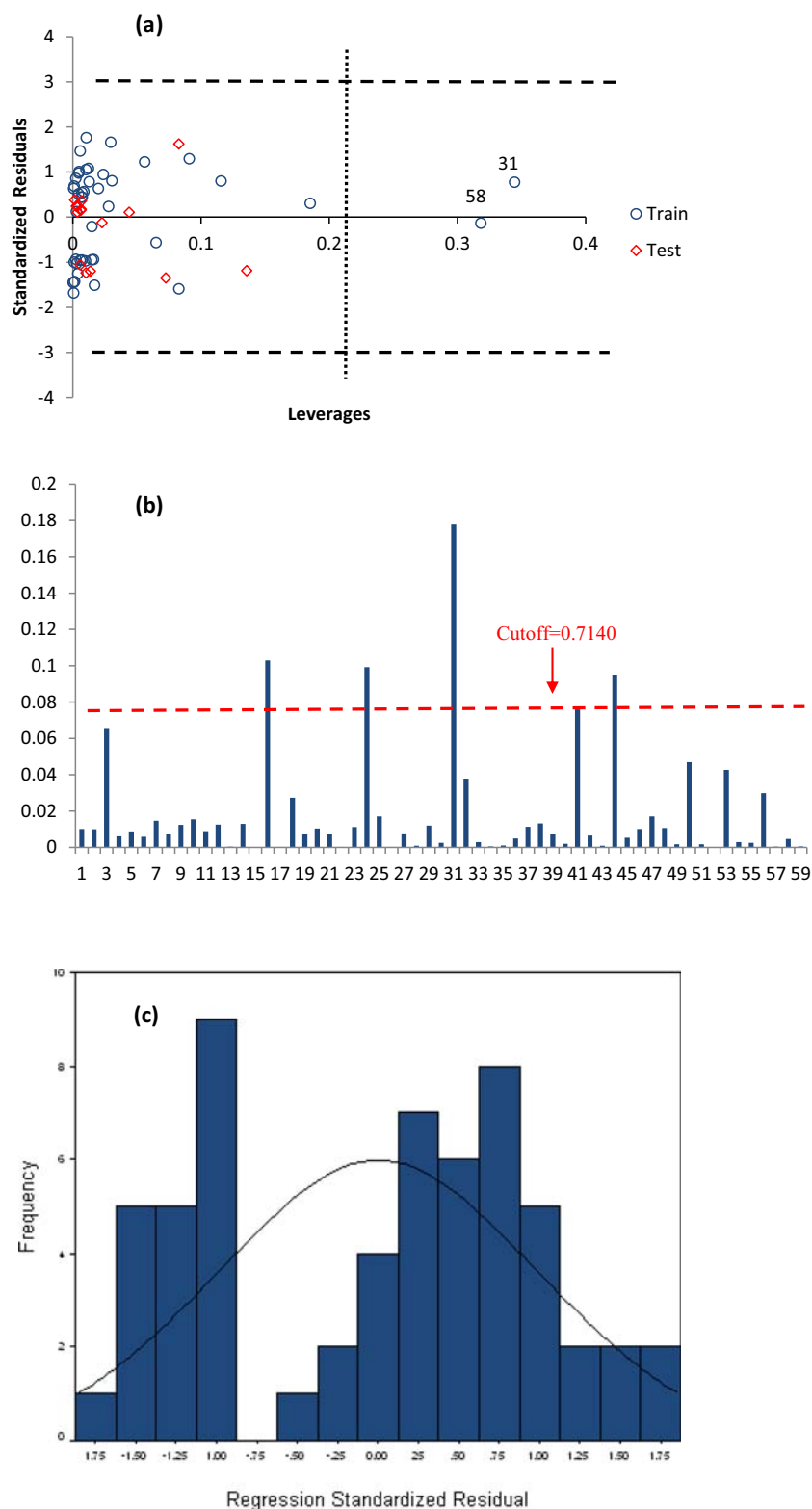
set compounds, and X^T is the transpose of X.

The warning leverage value (h^*), as a prediction tool, is expressed as:

$$h^* = \frac{3(k+1)}{n}$$

where k is the number of model descriptors and n is the number of training compounds.

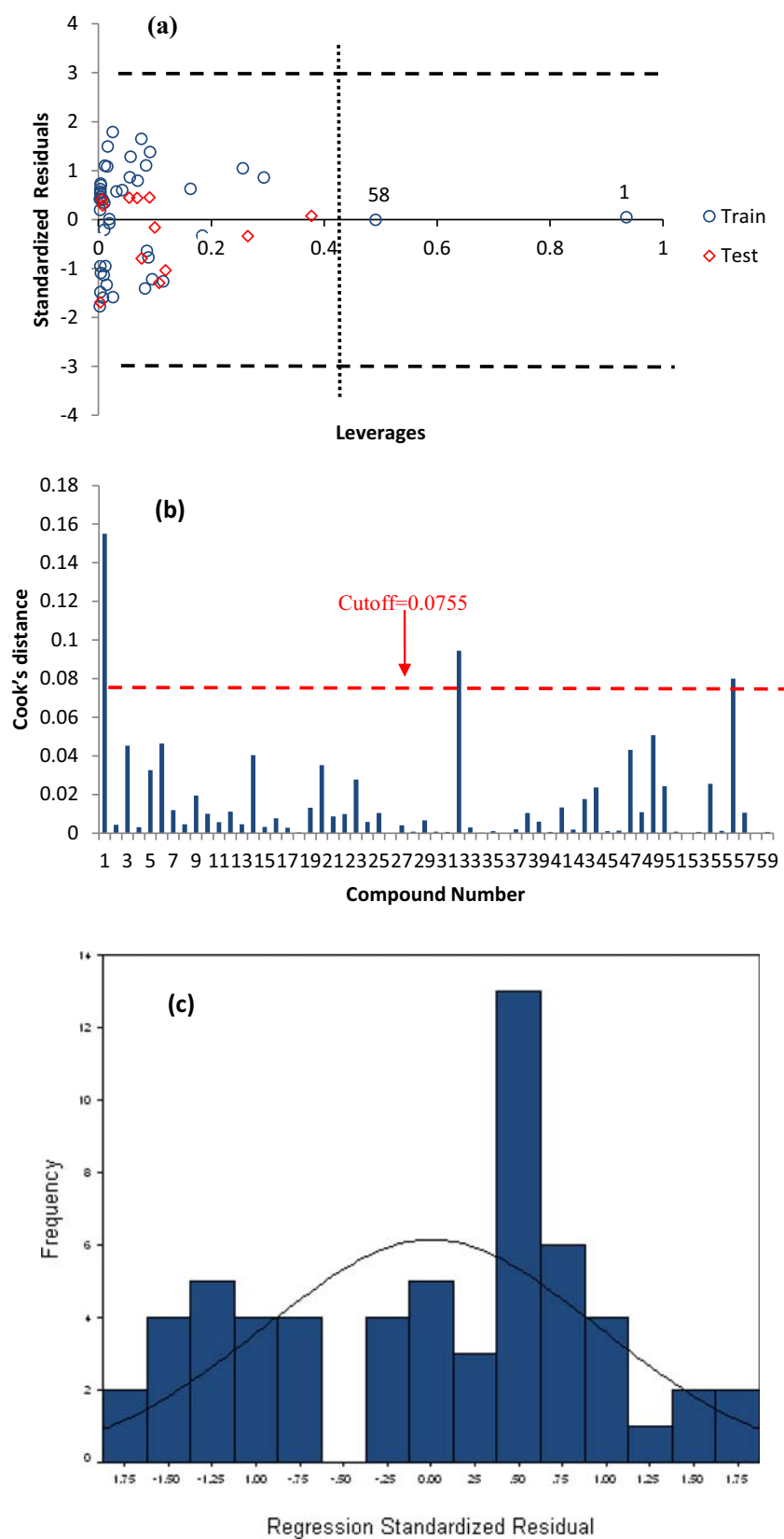
Fig. 8 Williams plot describing the applicability domain of the CoMFA model for the training and test sets ($h^* = 0.206$) (a); Cook's distance plot (b); and Histogram of model CoMFA residuals (c)



The Williams plot illustrates the distribution of data and its restricting range termed cutoff lines which all data should be between ± 3 units (horizontal dotted line) for standardized

residuals and the leverage value (h_i) should be less than warning leverage ($h_i < h^*$). The Williams plot for the training set is used to identify molecules with the greatest structural

Fig. 9 Williams plot describing the applicability domain of the CoMSIA model for the training and test sets ($h^* = 0.410$) (a), Cook's distance plot (b); and Histogram of model CoMSIA residuals (c)



influence ($h_i < h^*$) in developing the QSAR models. Molecules with $h_i > h^*$ are evaluated to be unreliably predicted by the models due to substantial extrapolation.

Cook's distance is used to estimate the influence of a single observation of the model [64] and is defined by Eq. (13):

$$D_i = \frac{e_i^2}{p+1} * \frac{h_i}{1-h_i} \quad (13)$$

where e_i^2 is the standard residual of the i th compound, p is the number of descriptors, and h_i is the leverage value of the i th compound. The cutoff of the Cook's distance is defined as $\frac{4}{(n-p-1)}$, and the compounds with Cook's distance higher than the cutoff value are marked as highly influential points of the model.

In this work, for CoMFA, CoMFA-RF, and CoMSIA models, most of the compounds fall into their corresponding application domain. These results indicated that our QSAR models had achieved a reliable activity prediction for the compounds.

As shown in the Williams plot of CoMFA model for the data set (Fig. 7a), two compounds (15 and 31) of training set had greater value than the warning leverage (h^*) value of 0.206. These compounds had low standard residual value and could be considered as influential in fitting the model performance but not necessarily outlier to be deleted from the training set. The test compounds were within the applicability domain (AD), indicating that their predicted activity values were reliable. Also, at the Cook's plot of CoMFA model (Fig. 7b); only, there were highly influential two compounds for training and test set that may slightly distort the regression. In addition, the histogram of the residuals distribution was confirmed with histogram plot as shown in Fig. 7c.

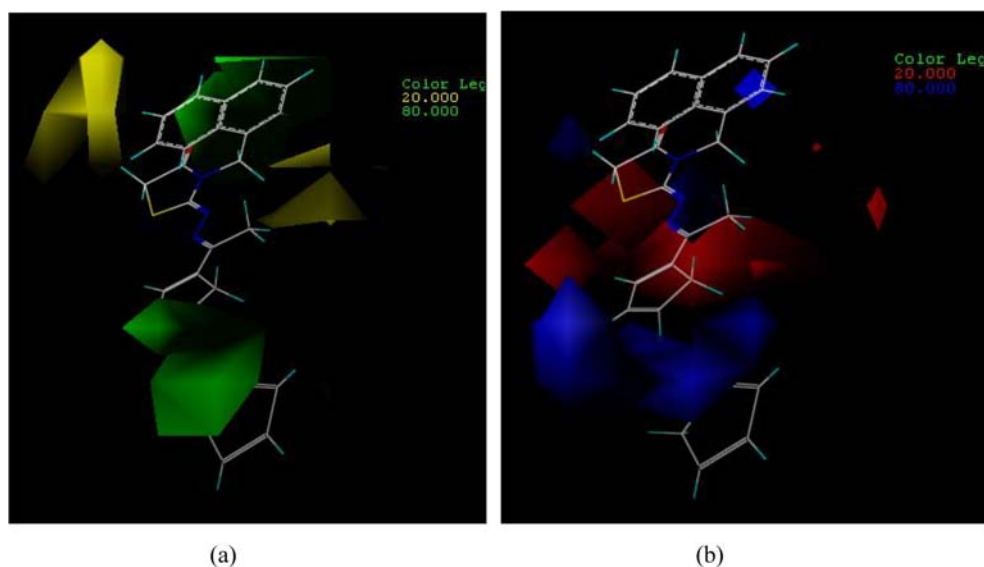
In the Williams plot of CoMFA-RF model for data set (Fig. 8a), two compounds (58 and 31) of training set had greater value than the warning leverage (h^*) value of 0.206. These compounds similar CoMFA model had low standard residual value and could be considered influential in fitting the model performance, but not necessarily outlier to be deleted from the training set. The test compounds were within the AD, indicating that their predicted activity values were reliable. Also, at the Cook's plot of CoMFA model (Fig. 8b), there were highly influential four compounds for training and test set. In addition, the histogram of the residuals distribution was confirmed with histogram plot as shown in Fig. 8c.

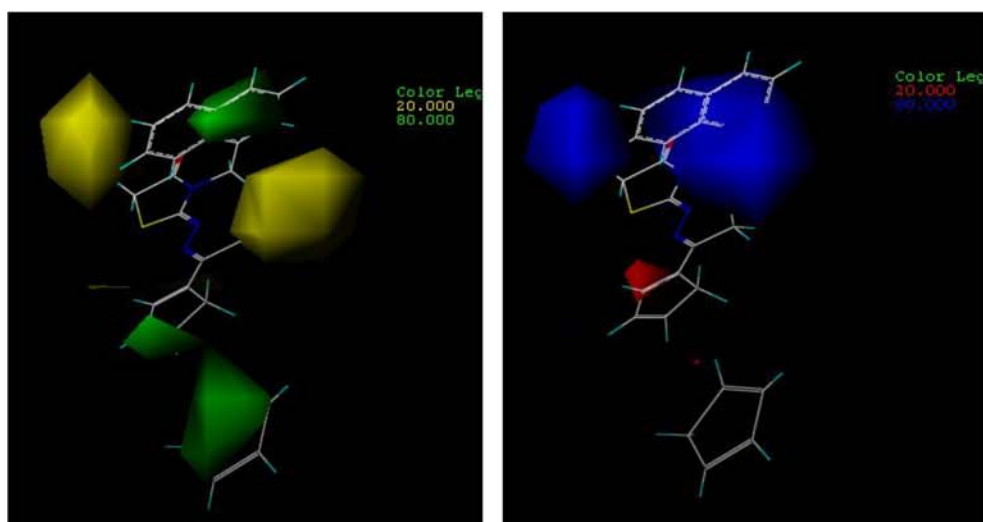
Also, at the Williams plot of CoMSIA model for data set (Fig. 9a), there were two outlier compounds for training set that could be regarded as structural outliers. Otherwise, according to the Cook's distances (cutoff = 0.0755) of the compounds in the data set, three highly influential compounds may distort the regression (Fig. 9b), also, the histogram of the residual distribution was confirmed with histogram plot as shown in Fig. 9c and prediction of CoMSIA model is reliable.

Interpretation of CoMFA and CoMSIA contour maps

The QSAR contour maps were used as an informative tool to visualize the effects of the different fields on the target compound 3D grid orientation of the models. The CoMFA and CoMSIA results were graphically interpreted by field contribution maps using the standard deviation (StDev) at each grid point and the coefficient from the PLS analysis (StDev*Coefficients).

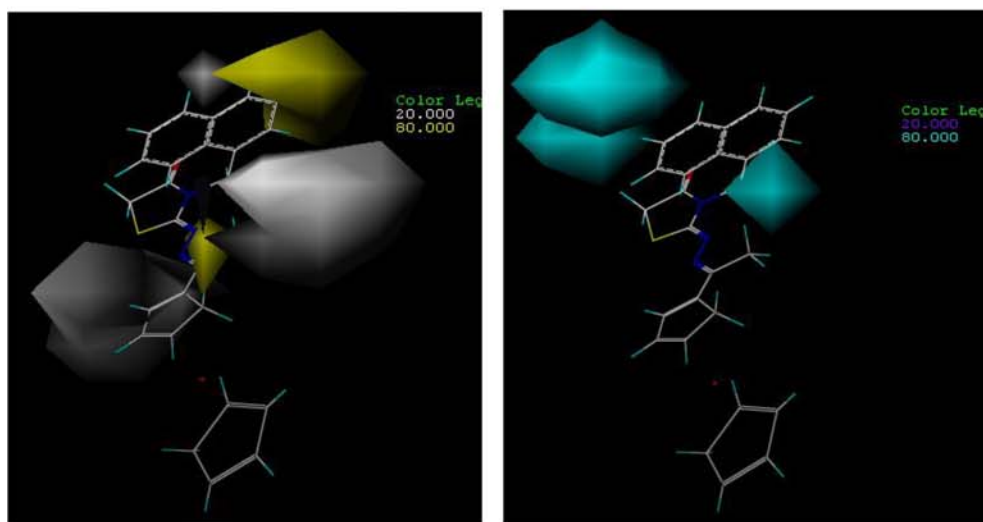
Fig. 10 CoMFAS $tDev$ *coeff. Contour plots with the combination of compound 36. **a** Steric contour maps: Green contours indicate regions where bulky groups increase activity and yellow contours indicate regions where bulky groups decrease activity. **b** Electrostatic contour maps: Blue contours indicate regions where positive charges increase activity and red contours indicate regions where negative charges increase activity





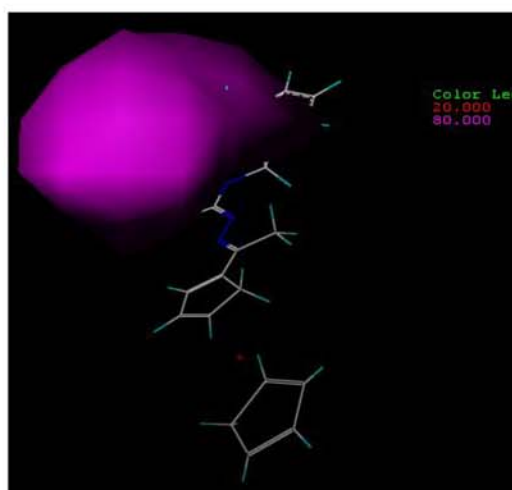
(a)

(b)



(c)

(d)



(e)

The CoMFA contour maps of the steric and electrostatic fields for the best anti-*T. gondii* agent (compound **56**) are shown in Fig. 10a, b. The field steric is shown by favorable groups (80% contribution) in green color and unfavorable ones (20% contribution) in yellow where the introduction of bulky groups may enhance or diminish the activity.

In the CoMFA steric maps, there was a green contour covering the naphthyl group at N-3 position of thiazolidin-4-one scaffold. The bulky groups at this position of compound improved anti-*T. gondii* activity and had the highest activity. Thiazolidin-4-ones substituted on the nitrogen atom of the 3-position with phenyl, naphthyl groups such as the compounds **13**, **15**, and **27–59** exhibited more potency, while compounds **2–5** due to the absence of these groups had relatively low activity. In addition, ferrocene group was substituted on the moiety aryl hydrazone that is attached to the carbon of the two positions of thiazolidin-4-one core with green contour in compound **56** had the highest activity. The compounds **1**, **13**, **15**, **17**, **22**, and **30–59** with bulky substituent's (e.g., aryl, thiophenyl, and butyl) at this region exhibited more potency.

Substituting the bulky groups at C-5 position of thiazolidin-4-one core, substituents of aryl moiety at N-3 position of thiazolidin-4-one core and the ethylidene group of hydrazone moiety decreased activity because these substituents were located at disfavored yellow contours. Therefore, these positions of thiazolidin-4-one core should be occupied by the steric moderate and low crowded substituents such as acetic acid and halogen aryl groups (e.g., **13**, **15**, **17**, **22**, **26**, **41**, **44**, **47**, and **53**).

In CoMFA electrostatic contour maps (Fig. 9b), the blue region (80% contribution) are favorable for electro-positive groups and red regions (20% contribution) is favorable for electronegative groups. The blue contour on the ferrocene group of compound **56** indicated the introduction of electropositive groups in this position could improve the biological activity. Besides, the red contours in the S-1 and C-5 positions of thiazolidin-4-

◀ **Fig. 11** CoMSIAStDev*Coeff contour plots with the combination of compound **56**. **a** *Steric contour maps*: green contours indicate regions where bulky groups increase activity; yellow contours indicate regions where bulky groups decrease activity; **b** *electrostatic contour maps*: blue contours indicate regions where positive charges increase activity; red contours indicate regions where negative charges increase activity. **c** *Hydrophobic contour maps*: yellow contours indicate regions where hydrophobic substituents enhance activity; white contours indicate regions where hydrophobic groups decrease activity. **d** *Hydrogen bond donor contour maps*: cyan contours indicate regions where H-bond donor groups increase activity and purple contours indicate the unfavorable regions for hydrogen bond donor substituents. **e** *H-bond acceptor contour maps*: magenta contours indicate regions where H-bond acceptor substituents increase activity; red contours indicate the disfavor regions for H-bond acceptor groups

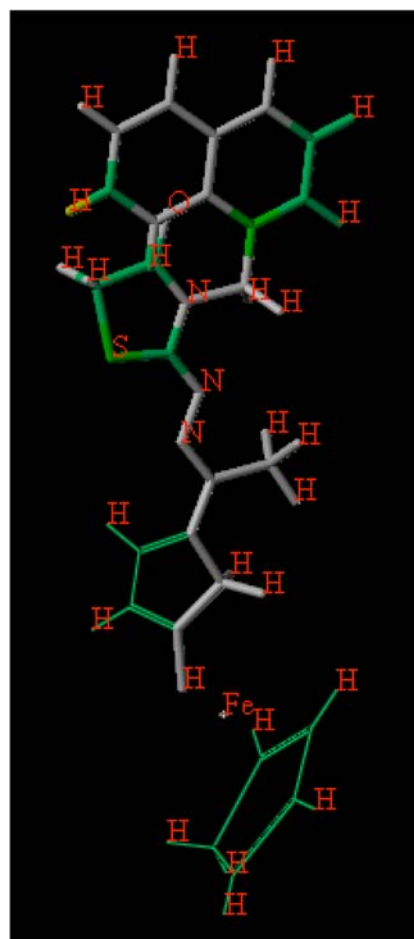


Fig. 12 The HQSAR contribution map of the most active compound (**56**). The colors in yellow, blue, green-blue, or green indicate positive contributions, while colors with red, red-orange, or orange represent negative contributions and intermediate contributions are colored in white

one core and N-1 of the moiety hydrazone showed that the electronegative substituent was beneficial to activity (compounds **25** < **24** < **23** < **26**, **22** < **13**, **15**, **17**).

In CoMSIA model, the steric and electrostatic, hydrophobic, hydrogen-bonding (H-bond) donor, and acceptor contour maps of compound **56** are shown in Fig. 11. The CoMSIA steric contours were nearly similar to that of CoMFA contours, so the electrostatic, hydrophobic interaction and hydrogen bond fields were described here. The blue contours located at the N-3 position of the thiazolidin-4-one core that increased anti-*T. gondii* activity (e.g. **5**, **6**, **23**, **54**, **55**, and **56**). The red contour was observed close to ethylidene hydrazone moiety that the introduction of the electronegative groups increased biological activity.

In the hydrophobic contour map, the yellow region is favorable (80% contribution) for the hydrophobic group while white region (20% contribution) is favorable for the hydrophilic group.

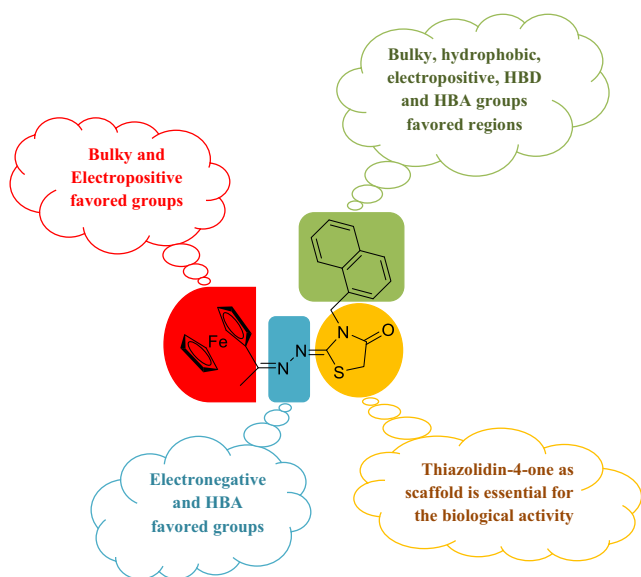


Fig. 13 Structure-activity relationship revealed by 3D- and 2D-QSAR and docking studies

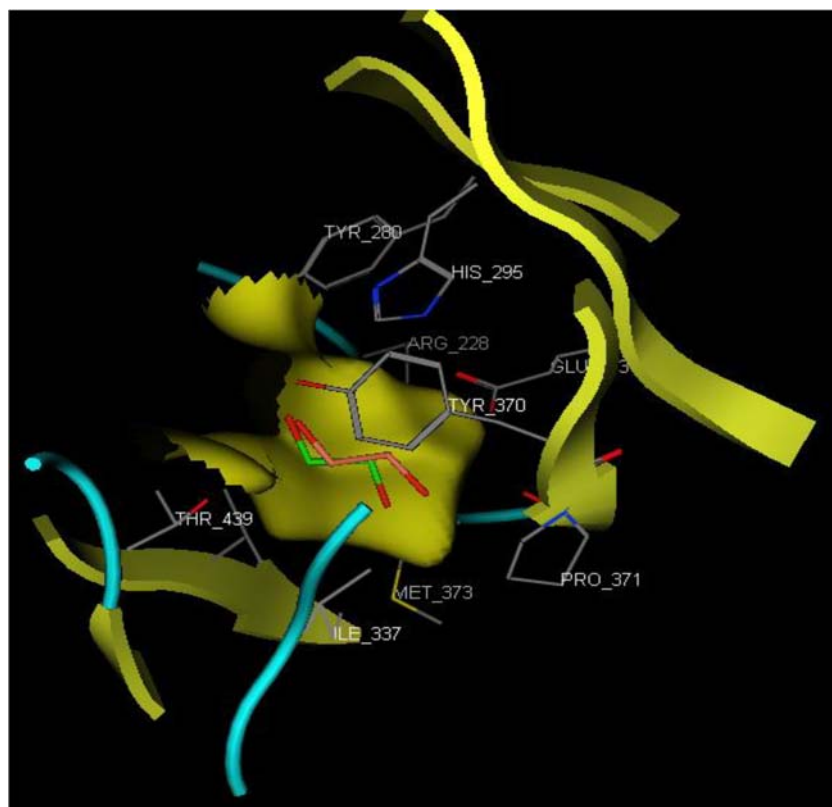
The white regions near the N-3 position of thiazolidin-4-one core and an aryl moiety of the ferrocene group showed that the introduction of hydrophilic groups into these positions might be beneficial for inhibitory activity (Fig. 11c). The yellow contours in a naphthyl ring of the N-3 position of thiazolidin-4-one

core and ethylidene hydrazineylidene moiety indicated that hydrophobic groups such as aryl and heterocyclic in this region could be increasing the activity of the compounds. The compounds **54–59** with hydrophobic substituent at this region exhibited more potency; while compounds **2–4** due to the absence of this hydrophobic group, had relatively lower activity. These results confirm that the yellow contour of hydrophobic map was in agreement with green contour of steric map.

The CoMSIA H-bond donor and acceptor contour maps correlated with hydrogen bond interactions of ligand with the target. The cyan and purple contour maps of H-bond donor indicated favorable (80% contribution) and unfavorable (20% contribution) interactions and the magenta and red contour maps indicated favorable (80% contribution) and unfavorable (20% contribution) H-bond acceptor groups (Fig. 11d, e). However, no unfavorable purple contour was observed. There were two cyan contours near to C-5 and C-6 positions of naphthyl ring and N-3 position of thiazolidin-4-one core that the H-bond donor groups might improve anti-*T. gonidii* activity.

Also, no unfavorable red contour for H-bond acceptor interaction was observed. There was a magenta contour in substituents of the N-3 position of thiazolidin-4-one core which was favorable for H-bond acceptor (compounds **57–59**) (Fig. 10e).

Fig. 14 Comparison of binding poses of co-crystallized ligand ethylene glycol (orange red) and its re-docked (green) in the active site of *TgROP8* (PDB: 3byv)



Interpretation of HQSAR contribution map

HQSAR calculations are based on the contributions of molecular fragments to the biological activity for each molecule. The results of the HQSAR contribution maps can be graphically shown as a color-coded structure diagram which the color of each atom reflects its contribution to the molecule's overall activity. The red end of the spectrum (red, red orange and orange) reflects negative contribution to the activity, while the green end (yellow, blue, green-blue, and green) represents positive effect and intermediate contributions are colored in white. The individual atomic contributions of the most active anti-*T. gondii* analogs (compound **56**) were displayed in Fig. 12.

The thiazolidin-4-one scaffold as maximal common structural fragment represented by green color code because it was a common fragment to all molecules and contributed in the same way to all inhibitors. The aryl derivatives in the N-3 position of thiazolidin-4-one core was highlighted in green and yellow colors, indicating the importance of these fragments to biological activity. Ferrocene group was substituted on the aryl hydrazone moiety was colored in green that positive contribution to inhibitory activity. Finally, the structure-activity relationship and binding features obtained by present QSAR models and molecular docking analysis are summarized in Fig. 13.

Molecular docking studies

TgROP8 is kinase from the rhoptyry organelles of the parasite that are unique to apicomplexan organisms. This protein contains a serine/threonine kinase domain is injected into the host cell in the precise moment of parasite internalization and manipulate the immune response of the host.

In this study, the MOE program was run to explore the possible binding modes of the anti-*T. gondii* agents. To confirm the validity of used docking parameters, the co-crystallized ligand ethylene glycol was re-docked into the active site of TgROP8 enzyme. The re-docking result and the cognate ligand (red) were almost completely superimposed and the RMSD value (0.8566 Å) guaranteed the reliability of the docking procedure (Fig. 14).

In order to gain functional and structural insight into the binding mode of the most-potent (compound **56**) and lest-potent (compound **10**) inhibitors and TgROP8 enzyme and also, to validate the results of QSAR contour maps, docking studies were carried out using MOE software (Fig. 15a, b).

Analysis of docking results revealed a high docking score (−12.44 kcal/mol) for the most active compound **56** in comparison with that of the least active compound **10** (−8.36 kcal/mol). The compound **10** bound with the less numbering of active site residues of enzyme and had less interaction with TgROP8 compared with compound **56**. The compound **56**

was well stabilized in the active site of TgROP8 and had significant interactions with the key amino acid residues of TgROP8 (Fig. S1 in the Supplementary file).

Regarding the docking studies, four types of interactions: hydrophobic π - π , π -cation, hydrogen-bonding, and hydrophobic interactions were involved in the attachment of compound **56** to the active site of the receptor. With few exceptions, the binding mode of the best-scored ligands with TgROP8 by LigX of MOE

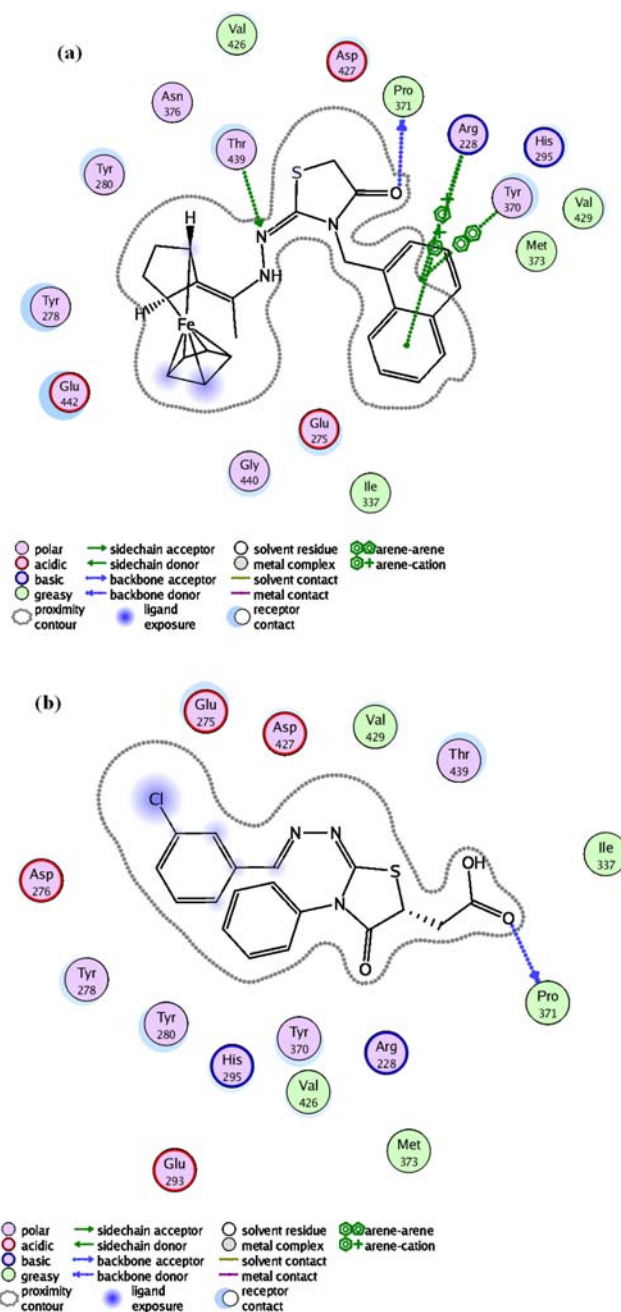


Fig. 15 **a** The 2D representation of the interaction between compound **56** (the most active compound). **b** The 2D representation of the interaction between compound **10** (the least active compound) in the crystal structure of TgROP8 (PDB ID: 3byv) using LigX in MOE

suggested that the compounds were oriented towards the gorge of the protein. In active site of the *TgROP8*, Arg 228 made two π -cation interactions with the naphthalene substituent. Also, Tyr 370 made π - π stacking interactions with naphthalene substituent.

The carbonyl oxygen in the thiazolidin-4-one core formed hydrogen bond with Pro371 while the nitrogen atom of hydrazine moiety could also form a hydrogen-binding interaction with Thr 439. Nevertheless, the ferrocene, naphthalene and thiazolidinone moieties made Van der Waals interactions with Glu 275, Ile 337, Gly 440, Glu 442, Tyr 278, Tyr 280, Asn 376, Val 426, Asp 427, His 295, Val 429, and Met 373. Therefore, with investigation of docking of all compounds at *TgROP8* active site, obtained results showed that the most potent compounds with Ferrocene and naphthalene moieties on thiazolidin-4-one core have more favorable interactions with *TgROP8* and bound with more number of active site residues in comparison with the less-potent compounds (Fig. S1 in the Supplementary file). These docking results validated the contour maps of QSAR models.

Molecular dynamics simulation

The molecular dynamics simulation has been done to elucidate behavior of *TgROP8* protein upon binding to the ligand and stability and interaction of ligand-protein throughout the simulation. The dynamics stability of secondary structure elements and conformational changes in protein-ligand complex were compared with protein of *TgROP8* by the root-mean-square derivations (RMSD) and radius of gyration (Rg) plots for 10 ns simulations with respect to temperature, potential energy, kinetic energy, and total energy, and results showed that this system is in a stable state.

The overall simulation convergence and ligand-protein equilibration were determined with RMSD of backbone atoms ($C\alpha$, C, and N) that is a measure of the stability

of the structures. The RMSD vs. Time is shown in Fig. 16.

This plot indicated that The RMSD of backbone (Ca, C, and N) of protein and protein-ligand complex (**56**) reached stability after about 4 and 2 ns of simulation, respectively, and RMSD value was under 2.0 Å in both the cases. The RMSD of protein-ligand complex (**56**) with an average of 1.51 ± 0.070 Å (mean \pm SD) was converged very close to protein with average RMSD value of 1.21 ± 0.112 Å was slightly larger than protein. It seemed that binding of ligand with protein increased the conformation flexibility of *TgROP8* protein.

Rg is a parameter that describes the equilibrium conformation of the native and the bound systems and is an indicator of the protein structure compactness.

Rg value as a function of time for protein and protein-ligand complex (**56**) were 18.59 ± 0.056 and 18.54 ± 0.100 Å, respectively. The graph showed that Rg of protein was increased for the first 2 ns of the simulation and then remained constant and Rg of protein-ligand complex (**56**) was constant after 1.50 ns of simulation and larger than protein As shown in Fig. 17.

Another simple way to measure the stability of ligand-protein complex is potential energy. The ligand bond protein potential energy was found to be $202,304.8 \pm 167.73$ kcal/mol indicating the stability of the system (Fig. 18).

The simulation results showed that the final structure and initial docked structure were in the same binding pocket and ligand-protein conformation was stable, and docking results were reliable. The 2D representation of the interaction between compound **56** after 10 ns simulation has been depicted in Fig. 19. This figure indicated that the interactions between most residues (Glu 275, Ile 337, Asn 376, Val 426, Tyr 278, Tyr 280, Arg 228, His 295, Glu 442, Met 373, Pro371, Thr 439, and Tyr 370) and compound **56** in the initial docked and final

Fig. 16 RMSD between *TgROP8* with ligand and without ligand

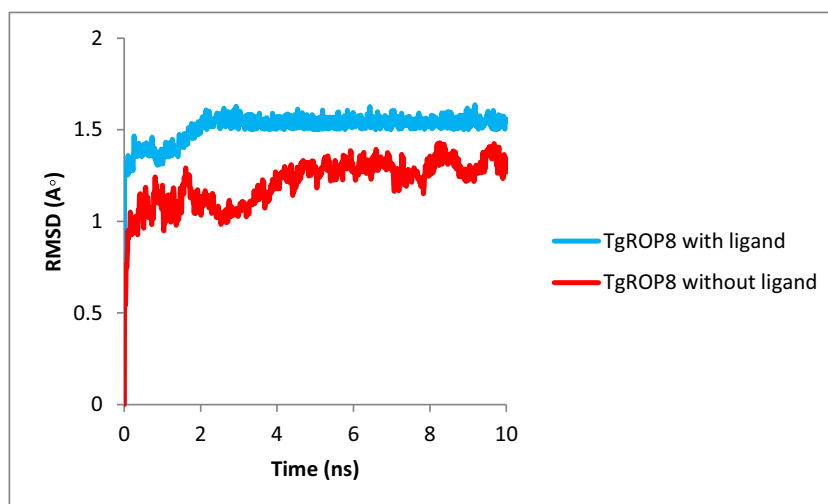


Fig. 17 R_g between TgROP8 with ligand and without ligand

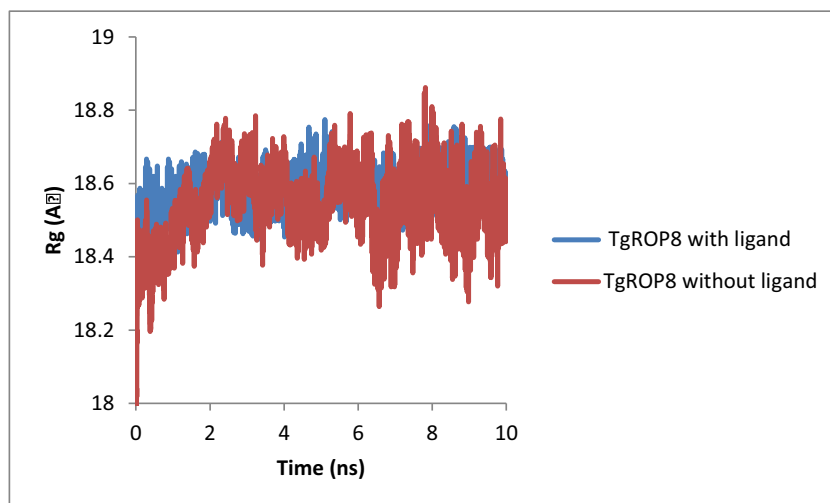


Fig. 18 Potential energy of TgROP8 with and without ligand

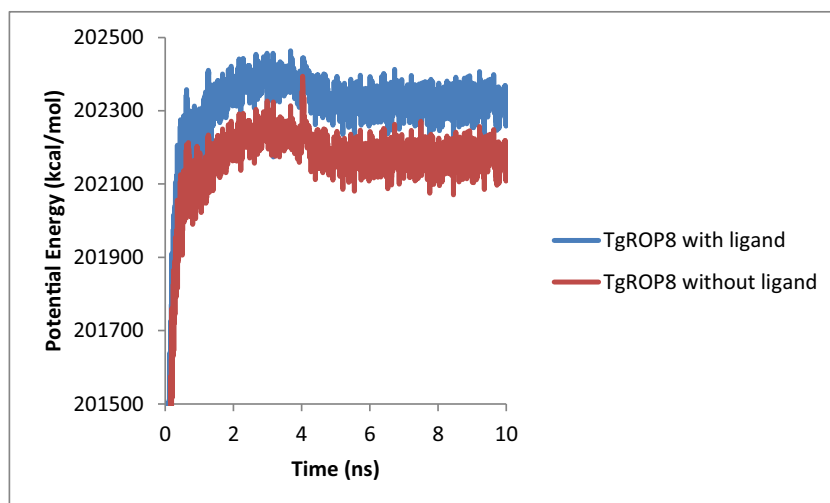
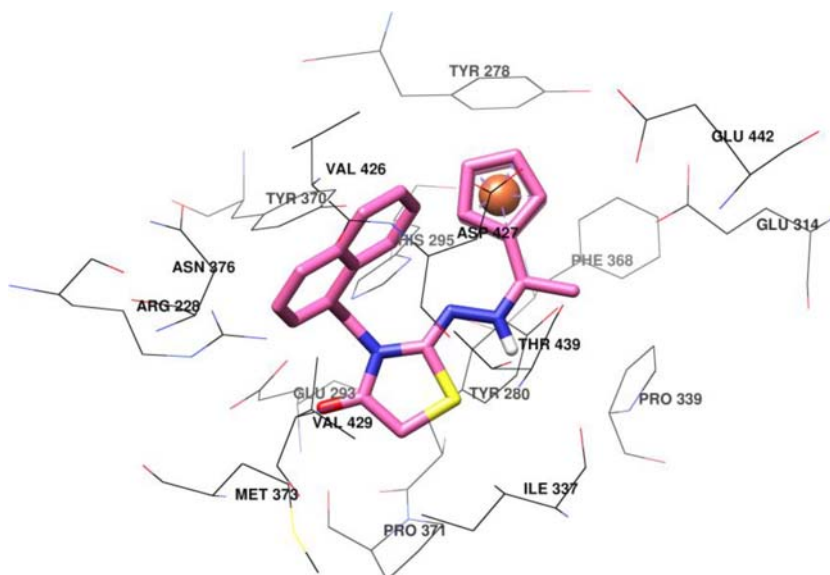


Fig. 19 3D representation of interactions between compound 56 and TgROP8 at the active site after 10 ns molecular dynamics simulations



protein-ligand complex (**56**) were unchanged. However, the number of amino acid residues at the active site had changed in this 3D representation. These binding interactions of compound **56** with Pro 371, Thr 439, Arg 228, and Tyr 370 similar to initial docking might be helpful to the stability of compound **56** in the active site of TgROP8 protein.

Conclusion

The 2D-(HQSAR) and 3D-QSAR (CoMFA, CoMFA-RF and CoMSIA) methods were employed to study a series of thiazolidin-4-one derivatives as anti-*T. gondii* agents. The CoMFA, CoMFA-RF, CoMSIA, and HQSAR models provided statistically significant results for internal and external validations including q^2 values of 0.897, 0.900, 0.910, and 0.924, r_{ncv}^2 values of 0.933, 0.935, 0.950, and 0.953, r_{pred}^2 values of 0.938, 0.998, 0.998, and 0.995, and r_m^2 values of 0.771, 0.725, 0.870, and 0.763, respectively. The CoMFA and CoMSIA contour maps and the HQSAR fragment contribution map were explained structure-activity relationship of this series of anti-*T. gondii* agents. Also, molecular docking and molecular dynamics simulation studies were carried out to confirm the rationality of the derived models. The thiazolidin-4-one core as scaffold and the bulky groups in the cyclic moieties as hydrophobic parts were key factors to improve inhibitory activity of TgROP8. These results showed good predictive models for the rational design of novel anti-*T. gondii* agents for the treatment of Toxoplasmosis disease.

Acknowledgments We are grateful to Clinical Biochemistry Research Center, Basic Health Sciences Institute, Sharekord University of Medicinal Sciences for financial support of this research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studied with human participants or animals performed by any of the authors.

References

- Murata Y, Sugi T, Weiss LM, Kato K (2017) Identification of compounds that suppress toxoplasma gondii tachyzoites and bradyzoites. PLoS One 12(6):e0178203
- Munera Lopez J, Ganuza A, Bogado SS, Munoz D, Ruiz DM, Sullivan Jr WJ, Vanagas L, Angel SO (2019) Evaluation of ATM kinase inhibitor KU-55933 as potential anti-toxoplasma gondii agent. Front Cell Infect Microbiol 9:26
- Zhang HB, Shen QK, Wang H, Jin C, Jin CM, Quan ZS (2018) Synthesis and evaluation of novel arctigenin derivatives as potential anti-toxoplasma gondii agents. Eur J Med Chem 158:414–427
- Johnson SM, Murphy RC, Geiger JA, DeRocher AE, Zhang Z, Ojo KK, Larson ET, Perera BG, Dale EJ, He P, Reid MC, Fox AM, Mueller NR, Merritt EA, Fan E, Parsons M, Van Voorhis WC, Maly DJ (2012) Development of toxoplasma gondii calcium-dependent protein kinase 1 (TgCDPK1) inhibitors with potent anti-toxoplasma activity. J Med Chem 55(5):2416–2426
- Wang ZD, Liu HH, Ma ZX, Ma HY, Li ZY, Yang ZB, Zhu XQ, Xu B, Wei F, Liu Q (2017) Toxoplasma gondii infection in immunocompromised patients: a systematic review and meta-analysis. Front Microbiol 8:389
- Singh S (2016) Congenital toxoplasmosis: clinical features, outcomes, treatment, and prevention. Trop Parasitol 6(2):113–122
- Hopper AT, Brockman A, Wise A, Gould J, Barks J, Radke JB, Sibley LD, Zou Y, Thomas S (2019) Discovery of selective toxoplasma gondii dihydrofolate reductase inhibitors for the treatment of toxoplasmosis. J Med Chem 62(3):1562–1576
- Sanford A, Schulze T, Potluri L, Hemsley R, Larson J, Judge A, Zach S, Wang X, Charman S, Vennerstrom JL (2018) Novel toxoplasma gondii inhibitor chemotypes. Parasitol Int 67(2):107–111
- de Aquino TM, Liesen AP, da Silva RE, Lima VT, Carvalho CS, de Faria AR, de Araujo JM, de Lima JG, Alves AJ, de Melo EJ (2008) Synthesis, anti-toxoplasma gondii and antimicrobial activities of benzaldehyde 4-phenyl-3-thiosemicarbazones and 2-[(phenylmethylene) hydrazono]-4-oxo-3-phenyl-5-thiazolidineacetic acids. Bioorg Med Chem 16(1):446–456
- Rocha-Roa C, Molina D, Cardona N (2018) A perspective on thiazolidinone scaffold development as a new therapeutic strategy for toxoplasmosis. Front Cell Infect Microbiol 8:360
- Foroutan M, Ghaffarifar F, Sharifi Z, Dalimi A, Jorjani O (2019) Rhopty antigens as toxoplasma gondii vaccine target. Clin Exp Vaccine Res 8(1):4–26
- Foroutan M, Ghaffarifar F, Sharifi Z, Dalimi A, Pirestani M (2018) Bioinformatics analysis of ROP8 protein to improve vaccine design against toxoplasma gondii. Infect Genet Evol 62:193–204
- Kubinyi H (1997) QSAR and 3D QSAR in drug design part 1: methodology. Drug Discov Today 2(11):457–467
- Kubinyi H (1997) QSAR and 3D QSAR in drug design. Part 2: applications and problems. Drug Discov Today 2(12):538–546
- Issar U, Arora R, Kumari T, Kakkar R (2019) Combined pharmacophore-guided 3D-QSAR, molecular docking, and virtual screening on bis-benzimidazoles and ter-benzimidazoles as DNA-topoisomerase I poisons. Struct Chem:1–17
- Akamatsu M (2002) Current state and perspectives of 3D-QSAR. Curr Top Med Chem 2(12):1381–1394
- Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design—a review. Curr Top Med Chem 10(1):95–115
- Gupta N, Vyas VK, Patel B, Ghate M (2014) Predictive 3D-QSAR and HQSAR model generation of isocitrate lyase (ICL) inhibitors by various alignment methods combined with docking study. Med Chem Res 23(6):2757–2768
- Floresta G, Amata E, Dichiaro M, Marrazzo A, Salerno L, Romeo G, Prezzavento O, Pittalà V, Rescifina A (2018) Identification of potentially potent heme oxygenase 1 inhibitors through 3D-QSAR coupled to scaffold-hopping analysis. ChemMedChem 13(13):1336–1342
- Fatima S, Gupta P, Agarwal SM (2018) Insight into structural requirements of antiamebic flavonoids: 3D-QSAR and G-QSAR studies. Chem Biol Drug Des 92(4):1743–1749
- Chavda J, Bhatt H (2019) 3D-QSAR (CoMFA, CoMSIA, HQSAR and topomer CoMFA) MD simulations and molecular docking studies on purinylpyridine derivatives as B-Raf inhibitors for the treatment of melanoma cancer. Struct Chem:1–15
- Wang F, Yang W, Shi Y, Le G (2015) 3D-QSAR, molecular docking and molecular dynamics studies of a series of ROR γ t inhibitors. J Biomol Struct Dyn 33(9):1929–1940

23. Ding L, Wang ZZ, Sun XD, Yang J, Ma CY, Li W, Liu HM (2017) 3D-QSAR (CoMFA, CoMSIA), molecular docking and molecular dynamics simulations study of 6-aryl-5-cyano-pyrimidine derivatives to explore the structure requirements of LSD1 inhibitors. *Bioorg Med Chem Lett* 27(15):3521–3528
24. Carvalho CS, Melo EJ, Tenorio RP, Goes AJ (2010) Anti-parasitic action and elimination of intracellular toxoplasma gondii in the presence of novel thiosemicarbazone and its 4-thiazolidinone derivatives. *Braz J Med Biol Res* 43(2):139–149
25. Liesen AP, de Aquino TM, Carvalho CS, Lima VT, de Araujo JM, de Lima JG, de Faria AR, de Melo EJ, Alves AJ, Alves EW, Alves AQ, Goes AJ (2010) Synthesis and evaluation of anti-toxoplasma gondii and antimicrobial activities of thiosemicarbazides, 4-thiazolidinones and 1,3,4-thiadiazoles. *Eur J Med Chem* 45(9):3685–3691
26. Carradori S, Secci D, Bizzarri B, Chimenti P, De Monte C, Guglielmi P, Campeste C, Rivanera D, Bordon C, Jones-Brando L (2017) Synthesis and biological evaluation of anti-toxoplasma gondii activity of a novel scaffold of thiazolidinone derivatives. *J Enzyme Inhib Med Chem* 32(1):746–758
27. Tenorio RP, Carvalho CS, Pessanha CS, de Lima JG, de Faria AR, Alves AJ, de Melo EJ, Goes AJ (2005) Synthesis of thiosemicarbazone and 4-thiazolidinone derivatives and their in vitro anti-toxoplasma gondii activity. *Bioorg Med Chem Lett* 15(10):2575–2578
28. Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose Tripos 5.2 force field. *J Comput Chem* 10(8):982–1012
29. Cramer III RD, Bunce JD, Patterson DE, Frank IE (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant Struct Act Relat* 7(1):18–25
30. Kellogg GE, Semus SF, Abraham DJ (1991) HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J Comput Aided Mol Des* 5(6):545–552
31. Saraiva APB, Miranda RM, Valente RPP, Araujo JO, Souza RNB, Costa CHS, Oliveira ARS, Almeida MO, Figueiredo AF, Ferreira JEV, Alves CN, Honorio KM (2018) Molecular description of alpha-keto-based inhibitors of cruzain with activity against Chagas disease combining 3D-QSAR studies and molecular dynamics. *Chem Biol Drug Des* 92(2):1475–1487
32. Borisa A, Bhatt H (2015) 3D-QSAR (CoMFA, CoMFA-RG, CoMSIA) and molecular docking study of thienopyrimidine and thienopyridine derivatives to explore structural requirements for aurora-B kinase inhibition. *Eur J Pharm Sci* 79:1–12
33. Sharma M, Jha P, Verma P, Chopra M (2019) Combined comparative molecular field analysis, comparative molecular similarity indices analysis, molecular docking and molecular dynamics studies of histone deacetylase 6 inhibitors. *Chem Biol Drug Des*
34. 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(24):4130–4146
35. Wang W, Tian Y, Wan Y, Gu S, Ju X, Luo X, Liu G (2019) Insights into the key structural features of N 1-aryl-benzimidazols as HIV-1 NNRTIs using molecular docking, molecular dynamics, 3D-QSAR, and pharmacophore modeling. *Struct Chem* 30(1):385–397
36. Ke YY, Shiao HY, Hsu YC, Chu CY, Wang WC, Lee YC, Lin WH, Chen CH, Hsu JT, Chang CW, Lin CW, Yeh TK, Chao YS, Coumar MS, Hsieh HP (2013) 3D-QSAR-assisted drug design: identification of a potent quinazoline-based Aurora kinase inhibitor. *ChemMedChem* 8(1):136–148
37. Dunn Iii W, Wold S, Edlund U, Hellberg S, Gasteiger J (1984) Multivariate structure-activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: the PLS method. *Quant Struct Act Relat* 3(4):131–137
38. Geladi P (1988) Notes on the history and nature of partial least squares (PLS) modelling. *J Chemom* 2(4):231–246
39. Kubinyi H, Martin YC, Folkers G (1993) 3D QSAR in drug design: volume 1: theory methods and applications, vol 1, Springer Science & Business Media
40. Bush BL, Nachbar Jr RB (1993) Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J Comput Aided Mol Des* 7(5):587–619
41. Zhao T, Zhao Z, Lu F, Chang S, Zhang J, Pang J, Tian Y (2019) Two-and three-dimensional QSAR studies on hURAT1 inhibitors with flexible linkers: topomer CoMFA and HQSAR. *Mol Divers*. <https://doi.org/10.1007/s11030-019-09936-5>
42. Wang F, Hu X, Zhou B (2019) Structural characterization of plasmodial aminopeptidase: a combined molecular docking and QSAR-based in silico approaches. *Mol Divers*. <https://doi.org/10.1007/s11030-019-09921-y>
43. Moda TL, Montanari CA, Andricopulo AD (2007) Hologram QSAR model for the prediction of human oral bioavailability. *Bioorg Med Chem* 15(24):7738–7745
44. Lowis DR (1997) HQSAR: a new, highly predictive QSAR technique. *Tripos Technical Notes* 1(5):17
45. Jain SV, Ghate M, Bhadoriya KS, Bari SB, Chaudhari A, Borse JS (2012) 2D, 3D-QSAR and docking studies of 1,2,3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. *Org Med Chem Lett* 2(1):22
46. Sainy J, Sharma R (2015) QSAR analysis of thiolactone derivatives using HQSAR, CoMFA and CoMSIA. *SAR QSAR Environ Res* 26(10):873–892
47. Zhang S, Lin Z, Pu Y, Zhang Y, Zhang L, Zuo Z (2017) Comparative QSAR studies using HQSAR, CoMFA, and CoMSIA methods on cyclic sulfone hydroxyethylamines as BACE1 inhibitors. *Comput Biol Chem* 67:38–47
48. Meng L, Feng K, Ren Y (2018) Molecular modelling studies of tricyclic triazinone analogues as potential PKC- θ inhibitors through combined QSAR, molecular docking and molecular dynamics simulations techniques. *J Taiwan Ins Chem E* 91:155–175
49. Stähle L, Wold S (1987) Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study. *J Chemom* 1(3):185–196
50. Alisi IO, Uzairu A, Abechi SE, Idris SO (2018) Evaluation of the antioxidant properties of curcumin derivatives by genetic function algorithm. *J Adv Res* 12:47–54
51. Tong J, Lei S, Qin S, Wang Y (2018) QSAR studies of TIBO derivatives as HIV-1 reverse transcriptase inhibitors using HQSAR, CoMFA and CoMSIA. *J Mol Struct* 1168:56–64
52. Wold S (1978) Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 20(4):397–405
53. Kearns M, Ron D (1999) Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural Comput* 11(6):1427–1453
54. Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Model* 20(4):269–276
55. Rácz A, Bajusz D, Héberger K (2015) Consistency of QSAR models: correct split of training and test sets, ranking of models and performance parameters. *SAR QSAR Environ Res* 26(7–9):683–700
56. Wang Z, Cheng L, Kai Z, Wu F, Liu Z, Cai M (2014) Molecular modeling studies of atorvastatin analogues as HMGR inhibitors using 3D-QSAR, molecular docking and molecular dynamics simulations. *Bioorg Med Chem Lett* 24(16):3869–3876
57. Lorca M, Morales-Verdejo C, Vásquez-Velásquez D, Andrades-Lagos J, Campanini-Salinas J, Soto-Delgado J, Recabarren-Gajardo G, Mella J (2018) Structure-activity relationships based on 3D-QSAR CoMFA/CoMSIA and design of aryloxypropanolamine agonists with selectivity for the human β -adrenergic

- receptor and anti-obesity and anti-diabetic profiles. *Molecules* 23(5):1191
58. Rücker C, Rücker G, Meringer M (2007) Y-randomization and its variants in QSPR/QSAR. *J Chem Inf Mod* 47(6):2345–2357
 59. Shapiro S, Guggenheim B (1998) Inhibition of oral bacteria by phenolic compounds part 1. QSAR analysis using molecular connectivity. *Quant Struct Act Relat* 17:327–337
 60. Jaiswal M, Khadikar PU, Scozzafava A, Supurn CT (2004) Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorg Med Chem* 14:3283–3290
 61. Weaver S, Gleeson MP (2008) The importance of the domain of applicability in QSAR modeling. *J Mol Graph Mod* 26(8):1315–1326
 62. Kaneko H, Funatsu K (2014) Applicability domain based on ensemble learning in classification and regression analyses. *J Chem Inf Mod* 54(9):2469–2482
 63. Huang S, Feng K, Ren Y (2019) Molecular modelling studies of quinazolinone derivatives as MMP-13 inhibitors by QSAR, molecular docking and molecular dynamics simulations techniques. *MedChemComm* 10(1):101–115
 64. Ruslin R, Amelia R, Yamin Y, Megantara S, Wu C, Arba M (2019) 3D-QSAR, molecular docking, and dynamics simulation of quinazoline–phosphoramidate mustard conjugates as EGFR inhibitor. *J Appl Pharm Sci* 9(01):089–097

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.