J Mol Model (2007) 13: 99–109
DOI 10.1007/s00894-006-0124-0

ORIGINAL PAPER

Shagufta · Ashutosh Kumar · Gautam Panda · Mohammad Imran Siddiqi

CoMFA and CoMSIA 3D-QSAR analysis of diaryloxy-methano-phenanthrene derivatives as anti-tubercular agents

Received: 12 December 2005 / Accepted: 26 April 2006 / Published online: 21 June 2006

Abstract Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) based on three-dimensional quantitative structure–activity relationship (3D-QSAR) studies were conducted on a series (44 compounds) of diaryloxy-methano-phenanthrene derivatives as potent antitubercular agents. The best predictions were obtained with a CoMFA standard model (q^2=0.625, r^2=0.994) and with CoMSIA combined steric, electrostatic, and hydrophobic fields (q^2=0.486, r^2=0.986). Both models were validated by a test set of seven compounds and gave satisfactory predictive r^2 values of 0.999 and 0.745, respectively. CoMFA and CoMSIA contour maps were used to analyze the structural features of the ligands to account for the activity in terms of positively contributing physicochemical properties: steric, electrostatic, and hydrophobic fields. The information obtained from CoMFA and CoMSIA 3-D contour maps can be used for further design of phenanthrene-based analogs as anti-TB agents. The resulting contour maps, produced by the best CoMFA and CoMSIA models, were used to identify the structural features relevant to the biological activity in this series of analogs. Further analysis of these interaction-field contour maps also showed a high level of internal consistency. This study suggests that introduction of bulky and highly electronegative groups on the basic amino side chain along with decreasing steric bulk and electronegativity on the phenanthrene ring might be suitable for designing better antitubercular agents.

Keywords Antitubercular agents · Diaryloxy-methano-phenanthrene · 3D-QSAR · CoMFA · CoMSIA

Shagufta · G. Panda
Medicinal & Process Chemistry Division,
Central Drug Research Institute,
Lucknow 226001, India

A. Kumar · M. I. Siddiqi (✉)
Division of Molecular & Structural Biology,
Central Drug Research Institute,
Lucknow 226001, India
e-mail: imsiddiqi@yahoo.com
Tel.: +91-522-2212412

Introduction

Tuberculosis (TB) kills more than 3 million people each year worldwide and, thus, remains one of the most deadly infectious diseases in the world. Annually, approximately ten million people develop the disease, with five million of these progressing to the infectious stage and, ultimately, three million dying from it. Despite having modern and sophisticated methods of prevention, early detection, diagnosis, and, thereafter, treatment for TB patients, the appearance of multi-drug-resistant (MDR) strains, with the threat of global human immunodeficiency virus, has led to tuberculosis being declared a "global emergency" by the World Health Organization [1]. Resistance has surfaced for all first-line drugs (isoniazid, rifampin, pyrazinamide, ethambutol, and streptomycin) and for several second-line and newer drugs (ethionamide, fluoroquinolones, macrolides, nitroimidazopyrans) [2, 3]. Because of this, there is an urgent need for anti-TB drugs with improved properties, such as enhanced activity against MDR strains, reduced toxicity, shortened duration of therapy, rapid mycobactericidal mechanism of action, and the ability to penetrate host cells and exert antitubercular effects in the intracellular environment [4]. Our continuing effort toward the design and development of new chemical entities for tuberculosis provided diaryloxy-methano-phenanthrenes as an encouraging lead for anti-TB agents [5]. Incorporation of basic aminoalkyl and 2-hydroxy-aminoalkyl moieties on the diaryloxy-methano-phenanthrene pharmacophore gave better antitubercular activity. To rationalize the design and, thus, to understand physiochemical properties and structural parameters of the pharmacophore, a three-dimensional quantitative structure activity relationship (3D-QSAR) study of diaryloxy-methano-phenanthrene derivatives by comparative molecular field analysis (CoMFA) [6] and comparative molecular similarity indices analysis (CoMSIA) [7] was performed. Traditional QSAR models are unable to explain complex structure–activity data because the extreme specificity of biological activity is described by 3-D intermolecular forces and predicated on 3-D molecular structures. Consequently, the most relevant

QSAR model would be shape-dependent and would describe steric and electrostatic interactions with sufficient accuracy. The comparative molecular field analysis (CoMFA) method meets these requirements. From its advent, CoMFA has been a powerful tool for studying 3D-QSAR. CoMFA examines differences in targeted properties that are related to changes in the shape of the steric and electrostatic fields surrounding the molecules. A QSAR table is used to accommodate the details of the shape of each field by sampling their magnitudes at regular intervals throughout a specified region of space [6]. More recently, another alternative molecular field analysis, CoMSIA, based on molecular similarity indices, was reported. This is an extension of the CoMFA methodology and differs only in the implementation of the fields. In CoMSIA, five different similarity fields are calculated: steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor. Similarity indices are calculated at regularly spaced grid points for the prealigned molecules. Instead of the direct measurement of the similarity between all mutual pairs of a molecule, indirect evaluation of the similarity of each molecule in the data set with a common probe atom is calculated [7]. A linear regression equation of similarity with biological activities is derived. This indirect ligand-based approach can assist in understanding structure–activity relationships (SARs) and can also serve as a tool in designing more potent antitubercular agents.

Computational methods

Data sets

A total of 44 diaryloxy-methano-phenanthrene based inhibitors with structures and activities were used in the study (Table 1) [5, 8].[1] The minimum inhibitory concentration (MIC) values were converted to the corresponding pMIC (−logMIC) and used as dependent variables in CoMFA and CoMSIA analyses. The pMIC values span a range of 3 log units, providing a broad and homogenous data set for the 3D-QSAR study. The initial structures of 44 compounds were constructed using the Insight II Builder module [9]. Conformations of compounds in the training set and test set were generated using the multisearch utility in Sybyl 7.0 (TRIPOS, 1699 South Hanley Road, St. Louis, MO 63144, USA). The conformer with the lowest energy was extracted and energy minimization was performed using the Tripos force field [10], with a distance-dependent dielectric, and the Powell conjugate gradient algorithm, with a convergence criterion of 0.01 kcal mol^{-1} Å. Thirty-seven compounds were used as training set and seven in the test set. The test set compounds were selected manually from each subgroup of the training set such that structural diversity and broad range of activity in the data set were included. The test set molecules were selected by considering the fact that this set of molecules represents a range of biological activity similar to that of the training set. Thus, the test set is the true representative of the training set. The MIC values were converted to pMIC (−log MIC) values and used as dependent variables in the CoMFA and CoMSIA calculations.

Molecular alignment

Structural alignment is one of the most sensitive parameters in 3D-QSAR analyses. The accuracy of the prediction of CoMFA and CoMSIA models and the reliability of the contour models depend strongly on the structural alignment of the molecules [11]. The molecular alignment was achieved by the Sybyl 7.0 routine database align. The most active compound 28 (Fig. 1a) was used as the alignment template, and the rest of the molecules were aligned on it by using the common substructure (Fig. 1b).

Calculations of atomic charges

Six different kinds of partial atomic charge were considered: (1) Gasteiger–Hückel charges [12], (2) Gasteiger–Marsili charges [13, 14], (3) MMFF94 charges [15], (4) Del Re charges [16], (5) Pullman charges [17], and (6) Hückel charges [18]. Both the Gasteiger–Marsili and Gasteiger–Hückel methods calculate atomic charges based on the information of the atoms and the connectivity within the molecule. The MMFF94 atomic charges are simply calculated based on the bond increment parameters in the MMFF94 force field. The calculation of Gasteiger–Marsili, Gasteiger–Hückel, MMFF94, Del Re, Pullman, and Hückel charges in SYBYL 7.0 was automated with SPL scripts.

CoMFA studies

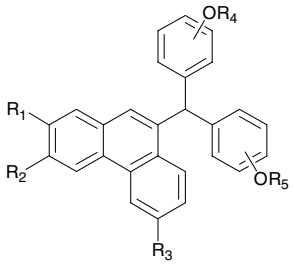
Steric and electrostatic CoMFA fields were calculated using the Lennard–Jones and the Coloumb potentials [6]. Default parameters (Tripos force field, dielectric distance $1/r^2$, steric and electrostatic cutoff 30 kal mol^{-1}, positively charged sp^3 hybridized carbon atom, grid spacing 2 Å) were used unless stated otherwise. With standard options for scaling of variables, the regression analysis was carried out using the full cross-validated partial least squares (PLS) method (leave one out) [19].

CoMSIA studies

The five similarity indices in CoMSIA, that is steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor descriptors, were calculated using a C^{1+} probe atom with a radius of 1.0 Å placed at a regular grid spacing

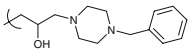
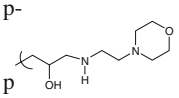
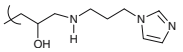
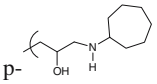
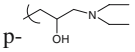
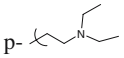
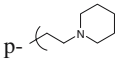
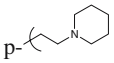
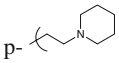
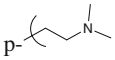
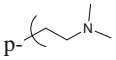
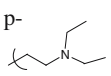
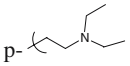
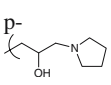
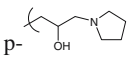
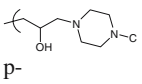
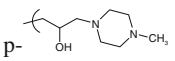
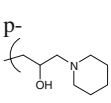
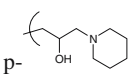
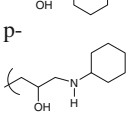
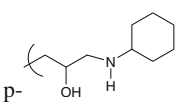
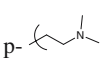
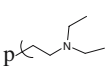
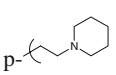
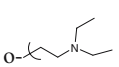
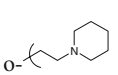
[1] Synthesis and antitubercular activity of compound 1–38 is reported in Reference [5a,b]. In vitro antitubercular activity of compounds 28, 29–31, and 39–44 (unpublished results).

Table 1 Structures and inhibitory activities of diaryloxy-methano-phenanthrene analogues



Comp. No.	R_1	R_2	R_3	R_4	R_5	MIC[a] (μg/mL)	-logMIC	CoMFA pMIC	Residual	CoMSIA pMIC	Residual
Training Set											
1	H,	H,	H	p-OCH_3	p-	6.25	4.905	4.911	-0.006	4.914	-0.009
2	H,	H,	H	p-OCH_3	p-	6.25	4.892	4.870	0.022	4.835	0.057
4	H,	H,	H	p-OCH_3	p-	6.25	4.894	4.914	-0.020	4.877	0.017
5	H,	H,	H	m-OCH_3	p-	12.5	4.603	4.598	0.005	4.577	0.026
6	H,	H,	H	m-OCH_3	p-	12.5	4.591	4.585	0.006	4.625	-0.034
7	H,	H,	H	m-OCH_3	p-	12.5	4.593	4.551	0.042	4.598	-0.005
9	H,	H,	H	o-OCH_3	p-	12.5	4.591	4.612	-0.021	4.624	-0.033
10	H,	H,	H	o-OCH_3	p-	12.5	4.567	4.569	-0.002	4.596	-0.029
11	H,	H,	H	o-OCH_3	p-	12.5	4.593	4.602	-0.009	4.626	-0.033
12	H,	H,	H	p-OCH_3	o-	12.5	4.603	4.592	0.011	4.623	-0.020
13	H,	H,	H	p-OCH_3	o-	12.5	4.591	4.600	-0.009	4.608	-0.017
14	H,	H,	H	p-OCH_3	o-	25	4.266	4.300	-0.034	4.302	-0.036
15	H,	H,	H	p-OCH_3	o-	6.25	4.894	4.896	-0.002	4.818	0.076
16	H,	H,	H	p-OCH_3	p-	12.5	4.629	4.643	-0.014	4.602	0.027
18	H,	H,	H	p-OCH_3	p-	6.25	4.941	4.947	-0.006	5.000	-0.059
19	H,	H,	H	p-OCH_3	p-	25	4.304	4.318	-0.014	4.329	-0.025
20	H,	H,	H	p-OCH_3	p-	25	4.329	4.290	0.039	4.298	0.031
21	H,	H,	H	p-OCH_3	p-	12.5	4.641	4.625	0.016	4.624	0.017

Table 1 (continued)

22	H,	H,	H	p-OCH_3		25	4.396	4.390	0.006	4.360	0.036
23	H,	H,	H	p-OCH_3		6.25	4.965	4.945	0.020	4.951	0.014
24	H,	H,	H	p-OCH_3		3.12	5.263	5.231	0.032	5.294	-0.031
25	H,	H,	H	p-OCH_3		3.12	5.254	5.283	-0.030	5.263	-0.010
27	H,	H,	H	p-OCH_3		6.25	4.915	4.911	0.004	4.942	-0.027
30	H,	H,	H	p-OH		12.5	4.580	4.600	-0.020	4.618	-0.038
31	H,	H,	H	p-OH		12.5	4.591	4.622	-0.031	4.591	0.000
32	H,	H,	H			6.25	4.981	4.956	0.025	4.944	0.037
33	H,	H,	H			6.25	4.919	4.934	-0.015	4.890	0.029
34	H,	H,	H			6.25	4.964	4.951	0.013	4.960	0.004
35	H,	H,	H			12.5	4.703	4.700	0.003	4.688	0.015
36	H,	H,	H			12.5	4.785	4.772	0.013	4.757	0.028
37	H,	H,	H			6.25	5.023	5.020	0.003	4.990	0.033
38	H,	H,	H			6.25	5.041	5.059	-0.018	5.048	-0.007
39	OCH_3	OCH_3	OCH_3	p-OCH_3		12.5	4.645	4.635	0.010	4.660	-0.015
40	OCH_3	OCH_3	OCH_3	p-OCH_3		12.5	4.666	4.690	-0.024	4.703	-0.037
41	OCH_3	OCH_3	OCH_3	p-OCH_3		25	4.374	4.394	-0.020	4.394	-0.020
43	OCH_3	OCH_3	OCH_3	p-OCH_3		25	4.365	4.345	0.020	4.355	0.010
44	OCH_3	OCH_3	OCH_3	p-OCH_3		25	4.374	4.370	0.004	4.347	0.027

of 2 Å. CoMSIA similarity indices $A^q_{F,K}(j)$ for a molecule j with atoms i at a grid point q are calculated by Eq. 1:

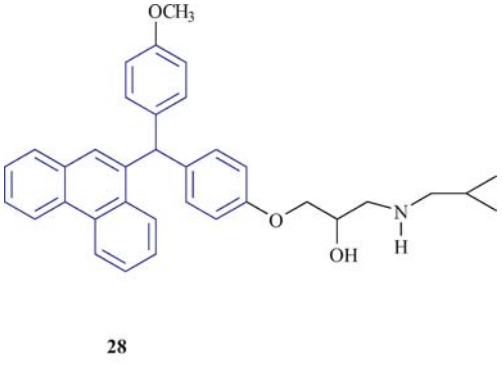
$$A^q_{F,K}(j) = \sum \omega_{probe,k}\omega_{ik}e^{-\alpha r^2_{iq}} \quad (1)$$

In this equation, k represents the following physiochemical properties: steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor. A Gaussian-type distance dependence was used between the grid point q and each

Table 1 (continued)

Test Set										
3	H,	H,	H	p-OCH_3	p-	12.5	4.567	4.572	-0.005	4.586	-0.019
8	H,	H,	H	o-OCH_3	p-	12.5	4.603	4.604	-0.001	4.650	-0.047
17	H,	H,	H	p-OCH_3	p-	12.5	4.617	4.618	-0.001	4.622	-0.005
26	H,	H,	H	p-OCH_3	p	6.25	4.920	4.921	-0.001	4.921	-0.001
28	H,	H,	H	p-OCH_3	p-	1.56	5.523	5.535	-0.012	5.099	0.424
29	H,	H,	H	p-OH	p-	12.5	4.554	4.560	-0.006	4.634	-0.080
42	OCH_3	OCH_3	OCH_3	p-OCH_3	p-	12.5	4.665	4.662	-0.003	4.659	0.006

[a]The abbreviation MIC stands for minimum inhibitory concentration, and it is defined as the concentration of the drug at which 90% of the mycobacteria ($H_{37}R_v$) was inhibited [5a–c]



28

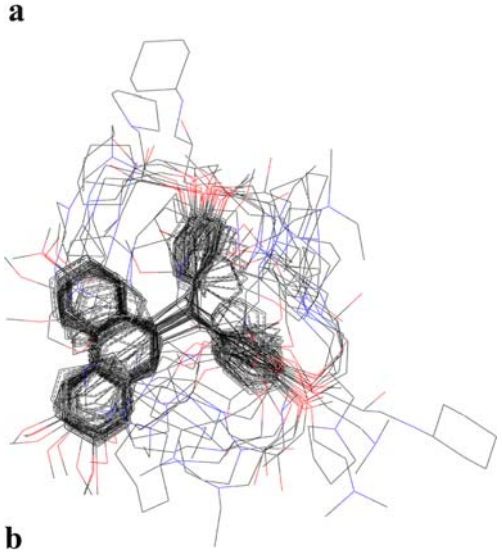


Fig. 1 **a** Template used for alignment (common substructure shown in *blue*). **b** Training set compounds aligned on minimum energy conformation of compound 28

atom *i* of the molecule. The default value of 0.3 was used as the attenuation factor (α). Here, steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from atomic partial charges, hydrophobic fields are derived from atom-based parameters [20], and H-bond and acceptor indices are obtained by a rule-based method based on experimental results [21].

Regression analysis

To derive 3D-QSAR models, the CoMFA and CoMSIA descriptors were used as independent variables and the pMIC as the dependent variable. Partial least squares (PLS) regression analysis was conducted with the standard implementation in the Sybyl 7.0 package. The predicted values of the models were evaluated by leave-one-out cross-validation. The cross-validated coefficient (q^2) was calculated using Eqs. 2 and 3 [22].

$$q^2 = 1 - \frac{PRESS}{\sum_{i=1}^{N} (y_i - y_m)^2} \tag{2}$$

$$PRESS = \sum_{i=1}^{N} \left(y_{pred_i} - y_i\right)_2 \tag{3}$$

where y_i is the activity for training set compounds, y_m is the mean observed value corresponding to the mean of the values for each cross-validation group, and $y_{pred,i}$ is the predicted activity for y_i.

Table 2 Influence of different partial charges on CoMFA models

	CoMFA1	CoMFA2	CoMFA3	CoMFA4	CoMFA5	CoMFA6
Partial charge	Gasteiger–Huckel	Gasteiger–Marsili	MMFF94	DelRe	Pullman	Huckel
q^{2a}	0.625	0.612	0.616	0.615	0.588	0.439
r^{2b}	0.994	0.993	0.983	0.988	0.983	0.920
S^{c}	0.021	0.024	0.036	0.031	0.036	0.079
F value[d]	1,088.387	840.517	364.612	509.249	361.521	71.769
N^{e}	5	5	5	5	5	5
Contributions						
Steric	0.391	0.407	0.449	0.415	0.439	0.684
Electrostatic	0.609	0.593	0.551	0.585	0.561	0.316
Column filtering	2.0					

[a]Cross-validated correlation coefficient
[b]Non-cross-validated correlation coefficient
[c]Standard error of estimate
[d]F test value
[e]Optimum number of components

Predictive r-squared

To validate the derived CoMFA and CoMSIA models, biological activities of an external test set of seven compounds (Table 1) were predicted using models derived from the training set. The predictive ability of the models is expressed by the predictive r^2 value, which is analogous to cross-validated r^2 (q^2) and is calculated using the formula

$$r^2_{pred} = \frac{SD - PRESS}{SD}$$

SD is the sum of the squared deviations between the biological activities of the test set molecules and PRESS is the sum of the squared deviations between the observed and the predicted activities of the test molecules.

Hardware and software

Insight II 2000.1 and Sybyl 7.0 were used for molecular modeling and 3D-QSAR analysis on a SGI Origin 300 workstation equipped with four 600-MHz R 12000 processors.

Results and discussion

CoMFA and CoMSIA 3D-QSAR models were derived using a series of derivatives based on diaryloxy-methano-phenanthrene, possessing antitubercular activities [8].[2] The chemical structures of the molecules and their actual pMIC values are shown in Table 1. The data set was divided into a training set of 37 and a test set of seven molecules (Table 1). For developing the CoMFA and CoMSIA models, all compounds of the training set were aligned on the minimum energy conformation of the most active molecule 28, obtained by using the Multisearch utility of Sybyl 7.0. The alignment of the training set is important as structural information with the nature of the receptor and its binding is not available.

CoMFA studies

The idea underlying the CoMFA methodology is that differences in biological activity are often related to differences in the magnitudes of molecular fields surrounding the receptor ligands investigated. The robustness and predictive ability of CoMFA models varies with the alignment, number of components, grid spacing, column filtering, and partial atomic charges. Proper prediction of the 3D-QSAR model depends on the method by which partial atomic charges are calculated in a particular data set. Therefore, to optimize the partial charges to be used, different sets of partial charges were used in building the CoMFA models and all of them exhibit good statistical quality. The statistical details are summarized in Table 2. The cross-validated correlation coefficient (q^2) for Gasteiger–Marsili, DelRe, MMFF94, and Pullman charges are more or less similar (0.612, 0.615, 0.616, and 0.588), but is highest in the case of the Gasteiger–Hückel charges (0.625) and lowest in the case of Hückel charges (0.439). The conventional non-cross-validated correlation coefficients (r^2) for Gasteiger–Hückel and Gasteiger–Marsili are comparable (0.994 and 0.993), and those for DelRe, MMFF94, and Pullman charges (0.988, 0.983 and 0.983) are also comparable, while it is lowest (0.920) in the case of Hückel partial charge. The standard error of estimation value for Gasteiger–Hückel partial charges is better than the other five partial charges. Based on the above observations, the best CoMFA model, obtained using Gasteiger–Hückel partial charges, was chosen for further analysis.

In CoMFA, the optimal number of components to be used in the analysis significantly influences the prediction

[2] Please see footnote 1.

Table 3 Influence of number of optimum number of components on the CoMFA models

	CoMFA1	CoMFA2	CoMFA3	CoMFA4	CoMFA5
Number	1	2	3	4	5
q^2	0.320	0.516	0.616	0.625	0.625
r^2	0.747	0.884	0.946	0.978	0.994
S	0.132	0.091	0.063	0.041	0.021
F value	103.290	129.074	191.430	357.849	1,088.387
Contributions					
Steric	0.425	0.444	0.415	0.397	0.391
Electrostatic	0.575	0.556	0.585	0.603	0.609

Table 4 Influence of column filtering value on the CoMFA models

	CoMFA1	CoMFA2	CoMFA3	CoMFA4	CoMFA5	CoMFA6
Column filtering (kcal mol^{-1})	0	1	1.5	2	2.5	3
q^2	0.606	0.617	0.623	0.625	0.605	0.619
r^2	0.995	0.995	0.996	0.994	0.983	0.980
S	0.020	0.019	0.019	0.021	0.037	0.040
F value	1,141.883	1,272.419	1,391.348	1088.387	348.174	299.863
Number	5	5	5	5	5	5
Contributions						
Steric	0.350	0.384	0.387	0.391	0.429	0.439
Electrostatic	0.650	0.616	0.613	0.609	0.571	0.561

ability of the model. The number of components describes the degree of complexity of the model; at some point adding more detail corresponds to fitting the data to noise, and the predictive ability begins to diminish. Usually, the optimal number of components is determined by selecting the highest q^2 value, which most often corresponds to the smallest S value. Whenever the last added component improved q^2 by less than about 5%, the less complex model was chosen. To find out the optimum number of components to be used in CoMFA studies, CoMFA models with different numbers of components were generated (Table 3). The best CoMFA model (q^2=0.625) and minimum standard error value (0.021) are obtained with five components, and further increase in the number of components has no effect on the q^2 value. Hence, five is selected as the optimum number of components for further analysis.

To explore the effect of column filtering on the CoMFA model with 2.0-Å grid spacing and Gasteiger–Hückel charges, different column-filtering values were used. Minimum sigma (Column filtering) values of 0, 1.0, 1.5, 2.5, and 3.0 kcal mol^{-1} (which were different from the default setting of 2.0 kcal mol^{-1}; Table 4) were also investigated, but all lead to a decrease in the q^2 value. The best q^2 was observed at 2.0 kcal mol^{-1}. Thus, to improve the signal-to-noise ratio by omitting those lattice points whose energy variation is below this threshold and also because of the computing time, the default setting at 2.0 kcal mol^{-1} was used in the further studies.

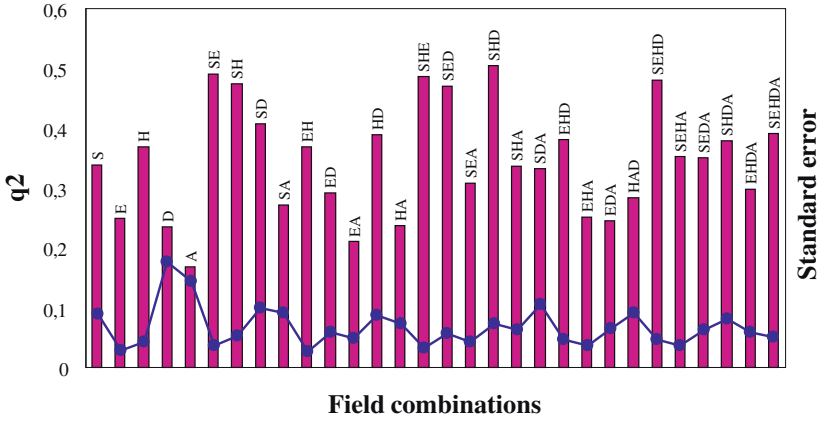


Fig. 2 Cross-validated r^2 (q^2) and standard error value for different field combinations. *S*, *E*, *H*, *D*, and *A* represent steric, electrostatic, hydrophobic, hydrogen bond donor, and acceptor fields, respectively

Table 5 Influence of column filtering value on the CoMSIA models

	CoMSIA1	CoMSIA2	CoMSIA3	CoMSIA4	CoMSIA5	CoMSIA6
Column filtering (kcal mol^{-1})	0	1	1.5	2	2.5	3
q^2	0.493	0.497	0.500	0.486	0.489	0.496
r^2	0.990	0.987	0.987	0.986	0.988	0.987
S	0.028	0.031	0.032	0.033	0.031	0.032
F value	597.257	479.485	460.408	427.948	503.461	465.969
Number	5	5	5	5	5	5
Contributions						
Steric	0.219	0.239	0.242	0.246	0.245	0.255
Electrostatic	0.520	0.487	0.482	0.474	0.475	0.466
Hydrophobic	0.262	0.274	0.276	0.280	0.279	0.279

CoMSIA studies

CoMSIA is an alternative approach to performing 3D-QSAR by comparative molecular field analysis. Molecular similarity is compared in terms of similarity indices. The CoMSIA method defines explicit hydrophobic (H) and hydrogen bond donor (D) and acceptor (A) descriptors in addition to the steric (S) and electrostatic fields (E) used in CoMFA. Primarily, the intention is to partition the different properties into various locations where they play a decisive role in determining the biological activity. In optimizing CoMSIA performance, the most important parameter is how to combine five fields in a CoMSIA model. To select the optimal result, we systematically changed the combination of field and PLS analysis of various CoMSIA models with different combinations of fields. Figure 2 shows the distribution of q^2 that resulted from the 30 field combinations. The CoMSIA model that included S, E, and H fields only performed relatively better than other field combinations. The q^2 of this model (0.486) is more or less equivalent to the SHD (Steric–Hydrophobic–Hydrogen bond donor) and SE (Steric–Electrostatic) model with q^2 of 0.504 and 0.490, respectively, but with a lower standard error value of 0.033 as compared to 0.072 and 0.037 of the SHD and SE models, respectively. Hence, the optimal field combination CoMSIA model that included the steric, electrostatic, and hydrophobic fields was chosen for further analysis.

Using the S, E, and H field combination, the sensitivity of the CoMSIA models to different column-filtering values was also investigated, and the results show that the column-filtering effect on the CoMSIA models agreed with the CoMFA models (Table 5). Here, the best q^2 was also obtained with the column-filtering value of 2.0 kcal mol^{-1}.

In CoMSIA, a Gaussian-type, distance-dependent function is used. In the preliminary parameter study, we calibrated the attenuation factor α to 0.3. To decide whether this is an appropriate value, α was varied in a parameter study within the range from 0.1 to 0.5 in steps of 0.1 and, subsequently, similarity indices and q^2 values were computed each time (Table 6). The q^2 value was highest (0.486) when the attenuation factor α was 0.3 with low standard error value (0.033). Thus, our study shows that α=0.3 is optimum for this data set.

Predictive power of CoMFA and CoMSIA models

To test the predictability of the analyses, the activities of training set compounds were calculated from the best CoMFA and CoMSIA models considering Gasteiger–Hückel partial charges; 2.0 kcal column filtering; five components; and steric, electrostatic, and hydrophobic field combinations. The correlations between experimental and the predicted activities for the training set for the CoMFA and CoMSIA models is shown in Fig. 3a,b,

Table 6 Attenuation factor

	CoMSIA1	CoMSIA2	CoMSIA3	CoMSIA4	CoMSIA5
α	0.1	0.2	0.3	0.4	0.5
q^2	0.321	0.486	0.486	0.479	0.482
r^2	0.947	0.973	0.986	0.991	0.993
S	0.064	0.046	0.033	0.026	0.023
F value	110.287	225.655	427.948	721.770	909.116
Contributions					
Steric	0.271	0.246	0.246	0.251	0.257
Electrostatic	0.536	0.512	0.474	0.441	0.414
Hydrophobic	0.193	0.242	0.280	0.308	0.330

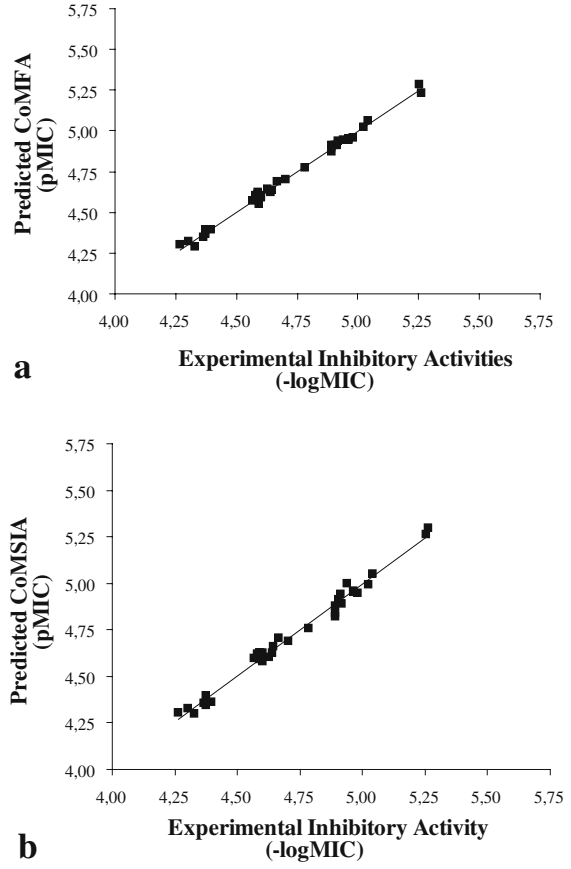


Fig. 3 **a** CoMFA and **b** CoMSIA predicted activities (pMIC) vs the experimental activities (−logMIC) of diaryloxy-methano-phenanthrene training set compounds

respectively, indicating good agreement of experimental and predicted values (Table 1). A set of seven manually selected compounds (3, 8, 17, 26, 28, 29, and 42 in Table 1), reserved as test set, was used to verify the efficacy of the CoMFA and CoMSIA models. The predictive correlation coefficient r^2 was 0.999 (CoMFA) and 0.745 (CoMSIA) (Table 1). Figure 4a,b shows the comparative plot of the predictions for the test set for the two models. In CoMFA, the predicted values fall very close to the actual pMIC, while the CoMSIA model predicted lower activity for one molecule of the test set. While both CoMFA and CoMSIA demonstrated good predictive ability, CoMFA is slightly better than CoMSIA.

CoMFA and CoMSIA contour maps

The CoMFA steric and electrostatic fields from the final non-cross-validated analysis were plotted as 3-D colored contour maps (Fig. 5). The field energies at each lattice point were calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table (SD*coeff), always plotted as the percentages of the contribution of CoMFA equation. These maps show regions where differences in molecular fields are associated with differences in biological activity. The CoMFA contours for steric and electrostatic fields are shown in Fig. 5, while those of CoMSIA steric, electrostatic, and hydrophobic fields are shown in Figs. 6, 7 and 8, respectively. In these contour maps, each colored contour represents particular properties such as green contours for regions of high steric tolerance (80% contribution), yellow for low steric tolerance (20% contribution), red contours for regions of decreased tolerance for positive charge (20% contribution), blue for regions of decreased tolerance for negative charge (80% contribution), yellow contours represent hydrophobically favored regions (80% contribution) and white contours for hydrophobically disfavored regions (20%) contribution). The larger size of the green–yellow region than the red–blue region indicates a greater contribution of steric fields than electrostatic ones in determining the biological activity.

As can be seen in Fig. 5, sterically favored large green polyhydra were found around the aminoalkyl and 2-hydroxy aminoalkyl groups attached to the *para* position of one of the phenyl rings, indicating that any bulkier substituent is preferred at this position for higher activity. The CoMSIA contour map for steric field (Fig. 6) also has a similar green polyhydron indicating the importance of steric bulk at the same region. Thus, this is an important

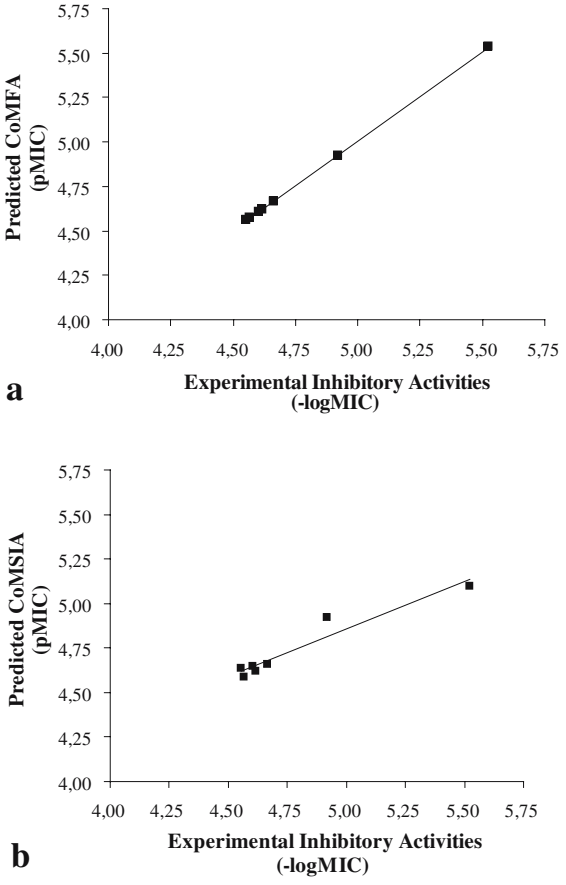


Fig. 4 **a** CoMFA and **b** CoMSIA predicted activities (pMIC) vs the experimental activities (−logMIC) of the testing set compounds

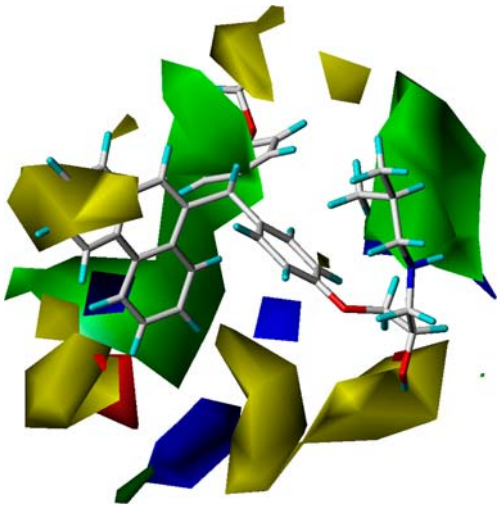


Fig. 5 Contour plot using CoMFA steric and electrostatic field

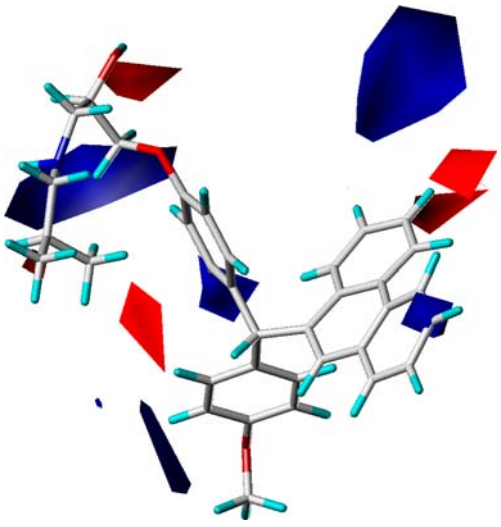


Fig. 7 Contour plot using CoMSIA electrostatic field

region as almost all contours show a similar-sized green polyhydron and indicate the importance of bulk in this region. However, this region is also important due to the presence of red and white polyhydra in CoMSIA contour plots in Figs. 7 and 8, respectively, indicating that increase of steric bulk, electronegative and hydrophilic groups in this region is favorable for the activity of this series of compounds (Table 1). Sterically non-favored yellow polyhydra; positively charged, favored blue regions; and hydrophobically favored yellow region are situated around the 1, 2, 3 positions of phenanthrene, indicating that the substitution of hydrogen by bulkier, electronegative, and hydrophilic groups like methoxy decreases the activity (compounds containing trimethoxy on phenanthrene).

Conclusions

In this study, 3-D CoMFA and CoMSIA QSAR analyses were used to predict the antitubercular activity of a set of diaryloxy-methano-phenanthrenes. The QSAR models gave good statistical results in terms of q^2 and r^2 values. The CoMFA model provided the most significant correlation of steric and electrostatic fields with the biological activities. Overall, the CoMFA method provided better statistical models than CoMSIA, which implies the significance of steric and electrostatic fields in the selectivity and activity of these compounds. The statistical significance and robustness of the 3D-QSAR models generated were confirmed using a test set. The effects of the steric, electrostatic, and hydrophobic fields around the aligned molecules on their activities were clarified by analyzing the CoMFA and CoMSIA contour maps. The information from this study suggests that incorporating bulk, higher degree of electronegativity, and hydrophilicity on basic amino side chain, along with diminishing steric bulk and electronegativity on phenanthrene nucleus, might be favorable for better antitubercular agents. It can be noted that a diaryloxy-methano-phenanthrene with the requisite groups could serve as a privileged structure for exploring antitubercular agents.

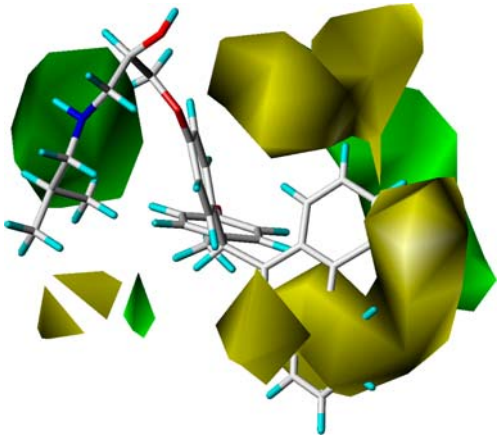


Fig. 6 Contour plot using CoMSIA steric field

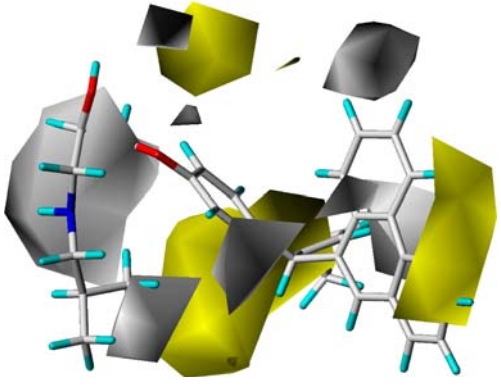


Fig. 8 Contour plot using CoMSIA hydrophobic field

Acknowledgements This manuscript is CDRI communication number 6893. This work was supported by the Council of Scientific and Industrial Research (CSIR) funded network project CMM0017—drug target development using in silico biology. Shagufta and Ashutosh thank CSIR for fellowships.

References

1. (a) Orme I, Secrist J, Anathan S, Kwong C, Maddry J, Reynolds R, Poffenberger A, Michael M, Miller L, Krahenbuh J, Adams L, Biswas A, Franzblau S, Rouse D, Winfield D, Brooks J (2001) Antimicrob Agents Chemother 45:1943–1946. (b) Baker RW, Mitscher LA, Arain TM, Shawer R, Stover CK (1996) Annu Rep Med Chem 31:161. (c) Dolin PJ, Raviglione MC, Kochi A (1994) Bull W H O 72:213–220. (d) Daffe M, Draper P (1998) Adv Microb Physiol 39:131–203. (e) Brennan PJ, Nikaido H (1995) Annu Rev Biochem 64:29–63. (f) Heifets L (1996) Antimicrob Agents Chemother 40:1759–1767
2. (a) Young DB, Duncan K (1995) Annu Rev Microbiol 49:641–673. (b) Schaeffer ML, Khoo KH, Besra GS, Chatterjee D, Brennan PJ, Belisle JT, Inamine JM (1999) J Biol Chem 274:31625–31631. (c) Collins L, Franzblau SG (1997) Antimicrob Agents Chemother 41:1004–1009. (d) Saito H, Tomioka H, Sato K, Emori M, Yamane T, Yamashita K, Hosoe K, Hidaka T, (1991) Antimicrob Agents Chemother 35:542–547
3. (a) Minnikin DE (1982) In: Ratedge C, Stanford J (eds) Chapter: The Biology of the Mycobacteria. Academic, San Diego, p 95. (b) Farmer P, Bayona J, Becerra M, Furin J, Henry C, Hiatt H, Kim JY, Mitnick C, Nardell E, Shin S (1998) Int J Tuberc Lung Dis 2:869–876. (c) Chopra I, Brennan P (1998) Tuber Lung Dis 78:89–98
4. (a) Teodori E, Dei S, Scapecchi S, Gualtieri F (2002) Farmaco 57:385–415. (b) Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C (2003) Lancet 362:887–899. (c) Smith CV, Sharma V, Sacchettini JC (2004) Tuberculosis (Edinb) 84:45–55
5. (a) Panda G, Shagufta, Mishra JK, Chaturvedi V, Srivastava AK, Srivastava R, Srivastava BS (2004) Bioorg Med Chem 12:5269–5276. (b) Panda G, Shagufta, Srivastava AK, Sinha S (2005) Bioorg Med Chem Lett 15:5222–5225. (c) Siddiqi S (1992) In: Clinical microbiology handbook, vol. 1. ASM, Washington DC
6. Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
7. Klebe G, Abraham UJ (1999) J Comput Aided Mol Des 13: 1–10
8. Synthesis of compound **39–44**: Shagufta, Srivastava AK, Sharma R, Mishra R, Balapure AK, Murthy PSR, Panda G (2006) Bioorg Med Chem 14(5):1497–1505
9. Insight II 2000.1 Program (2000) Accelrys, San Diego, CA
10. Clark MC, Cramer RD III, van Opden Bosch N (1989) J Comput Chem 10:982–1012
11. Cho SJ, Tropsha A (1995) J Med Chem 38:1060–1066
12. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219–3288
13. Marsili M, Gasteiger J (1980) Croat Chem Acta 53:601–614
14. Gasteiger J, Marsili M (1981) Org Magn Reson 15:353–360
15. Halgren TA (1996) J Comput Chem 17:490–519
16. Del Re G (1958) J Chem Soc 4031–4044
17. Berthod H, Pullman A (1965) J Chem Phys 42:942–946
18. Purcell WP, Singer JA (1967) J Chem Eng Data 12:235–246
19. Bush BL, Nachbar RB Jr (1993) J Comput Aided Mol Des 7:587–619
20. Viswanadhan VN, Ghose AK, Revenker GR, Robins R (1989) J Chem Inf Comput Sci 29:163–172
21. Klebe G (1994) J Mol Biol 237:212–235
22. Leach AR (2001) Molecular modelling: Principles and Applications. Henry Ling, London, p 695