ORIGINAL PAPER

# Receptor based 3D-QSAR to identify putative binders of *Mycobacterium tuberculosis* Enoyl acyl carrier protein reductase

**Ashutosh Kumar · Mohammad Imran Siddiqi**

**Abstract** In the current study, the applicability and scope of 3D-QSAR models (CoMFA and CoMSIA) to complement virtual screening using 3D pharmacophore and molecular docking is examined and applied to identify potential hits against *Mycobacterium tuberculosis* Enoyl acyl carrier protein reductase (MtENR). Initially CoMFA and CoMSIA models were developed using series of structurally related arylamides as MtENR inhibitors. Docking studies were employed to position the inhibitors into MtENR active site to derive receptor based 3D-QSAR models. Both CoMFA and CoMSIA yielded significant cross validated $q^2$ values of 0.663 and 0.639 and $r^2$ values of 0.989 and 0.963, respectively. The statistically significant models were validated by a test set of eight compounds with predictive $r^2$ value of 0.882 and 0.875 for CoMFA and CoMSIA. The contour maps from 3D-QSAR models in combination with docked binding structures help to better interpret the structure activity relationship. Integrated with CoMFA and CoMSIA predictive models structure based (3D-pharmacophore and molecular docking) virtual screening have been employed to explore potential hits against MtENR. A representative set of 20 compounds with high predicted $IC_{50}$ values were sorted out in the present study.

**Keywords** CoMFA · CoMSIA · 3D-Pharmacophore · Molecular docking · Virtual screening

A. Kumar · M. I. Siddiqi (✉)
Molecular and Structural Biology Division,
Central Drug Research Institute,
Lucknow 226 001, India
e-mail: imsiddiqi@yahoo.com
e-mail: mi_siddiqi@cdri.res.in

## Introduction

*Mycobacterium tuberculosis* contains unique signature fatty acids, the mycolic acids that are unusually long chain R-alkyl, β-hydroxy fatty acids of 60-90 carbons [1]. The TB-specific drugs isoniazid (isonicotinic acid hydrazide (INH)) and ethionamide have been shown to target the synthesis of these mycolic acids, which are central constituents of the mycobacterial cell wall. Among the enzymes involved in mycolic acid biosynthesis, the NADH-dependent enoyl acyl carrier protein reductase (MtENR) encoded by the Mycobacterium gene *inhA* is a key catalyst in mycolic acid biosynthesis. Studies over the years have established that MtENR is the primary molecular targets of INH [2], the drug that for the past 40 years has been, and continues to be, the frontline agent for the treatment of TB. As a prodrug, INH must first be activated by KatG, a catalase-peroxidase that oxidizes INH to an acyl radical that binds covalently to NADH, the cosubstrate for MtENR [3]. The INH-NADH adduct then functions as a potent inhibitor of MtENR. The requirement for INH activation opened a backdoor for the development of drug resistance by *M. tuberculosis*. Indeed, KatG-associated mutations account for 50% of the INH-resistant clinical isolates [4]. Thus, direct MtENR inhibitors that avoid this activation requirement would be promising candidates for the development of novel antitubercular agents. Recent studies focused on the development of inhibitors targeting MtENR directly without the requirement for activation. Several series of direct MtENR inhibitors including pyrazole derivatives, indole-5-amides [5], alkyl diphenyl ethers [6], pyrrolidine carboxamides [7] and arylamides [8] have been identified that show good inhibitory potency. Despite these and other known inhibitors, more structurally diverse inhibitors of MtENR need to be

discovered for improving the understanding of the biological function of MtENR and to discover its potential therapeutic indications. With a view to identify potential compounds with higher predicted potencies, in the present study, an attempt of ligand and structure-based rational searching of putative active site binders of MtENR was made using the available crystal structure of MtENR-Genz10850 [5] and series of arylamides with known inhibitory activity [8]. Such a method allows combining receptor-based and ligand-based approaches, thus utilizing most of the currently available structural data. The design of new compounds should never be based on one single approach. Although any single approach either ligand based approaches like QSAR or 3D pharmacophore or structure based approach like molecular docking can be used to design new compounds but better results can be achieved using the consensus of both ligand and structure based design approaches. The integrated ligand and receptor based approach has been successfully used by our group for the identification of *M. tuberculosis* Thymidine monophosphate kinase inhibitors as novel antitubercular lead compounds [9]. A good approach to design a putative library targeting a particular protein might be to combine 3D pharmacophore, molecular docking and scoring with QSAR by taking the top results identified separately by the different methods. So to identify putative MtENR binders first, CoMFA and CoMSIA studies were performed on a training set consisting of arylamides as MtENR inhibitors using molecular docking as alignment strategy. Second, the as-built CoMFA and CoMSIA models were validated by testing set. Finally, a ligand based pharmacophore was generated using crystal structure bound conformation of potent MtENR inhibitor Genz10850, which was then used in the virtual screening of the Maybridge screening collection. Twenty compounds with high screening scores obtained by jointly using a series of virtual screening methods were further supported by the CoMFA and CoMSIA models with high predicted $IC_{50}$ values.

## Materials and methods

### Dataset

*In vitro* inhibitory activity data ($IC_{50}$ (μM)) of the arylamides on MtENR, reported by Kuo *et al.* [5] and Xin He *et al.* [8] was taken for the study. Thirty seven molecules were selected for developing the model and the rest of the molecules whose $IC_{50}$ values reported as >100 were not considered in this study. The structures of the compounds and their biological data are given in Table 1. The $IC_{50}$ values were converted to the corresponding $pIC_{50}$

(-log $IC_{50}$) and used as dependent variables in CoMFA and CoMSIA analysis. The $pIC_{50}$ values span a range of 3-log units providing a broad and homogenous data set for 3D-QSAR study. The 3D-QSAR models were generated using a training set of 29 molecules. Predictive power of the resulting models was evaluated using a test set of eight molecules (Table 1 marked with *). The test compounds were selected randomly such that a wide range of activity in the data set was included. At physiological pH both the nitrogens of piperazine ring remain in unprotonated form, therefore all the molecules were modeled in neutral state.

### Molecular docking and alignment

FlexX [10] based molecular docking was carried out to get the appropriate binding conformations of the arylamides inhibitors and virtual screening hits into MtENR binding pocket. All the molecules were docked into the inhibitor binding site of MtENR crystal structure in complex with Genz-10850 (PDB entry code 1P44) [5].

### CoMFA studies

Steric and electrostatic interactions were calculated using the Tripos force field [11] with a distance-dependent dielectric constant at all interactions in a regularly spaced (2Å) grid taking C.3 carbon atom as steric probe and a single positive charge as electrostatic probe. Values of the steric and electrostatic fields were truncated at 30.0 kcal mol$^{-1}$. The minimum sigma (column filtering) was set to 2.0 kcal mol$^{-1}$ to improve the signal to noise ratio by omitting those lattice points whose energy variation was below this threshold. A partial least-squares (PLS) approach [12–14], which is an extension of multiple regression analysis, was used to derive the 3D QSAR models in which the CoMFA descriptors were used as independent variables, and the experimental $pIC_{50}$ values were used as dependent variables. The cross-validation with leave-one-out (LOO) option and the SAMPLS program [15] was applied to obtain the optimal number of components to be used in the final analysis. After the optimal number of components was determined, a non-cross-validated analysis was performed without column filtering.

### CoMSIA studies

In CoMSIA, a distance-dependent Gaussian-type physico-chemical property has been adopted to avoid singularities at the atomic positions and dramatic changes of potential energy for those grids in the proximity of the surface. With the standard parameters and no arbitrary cutoff limits, three physico-chemical properties, that is, steric, electrostatic and hydrophobic fields were calculated. The steric contribution

**Table 1** Structures and activities of arylamides used for developing CoMFA and CoMSIA models. Compounds marked as (*) belong to test set



| Cmpds | X | N | R1 | R2 | IC$_{50}$ (μM) | -logIC$_{50}$ | CoMFA pIC$_{50}$ | CoMSIA pIC$_{50}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | N | 0 | H | H | 38.86 | 4.41 | 4.326 | 4.469 |
| 2 | N | 0 | 4-CH$_3$ | H | 16.64 | 4.779 | 4.723 | 4.641 |
| 3 | N | 0 | 4-CH$_3$ | 3-CF$_3$ | 6.26 | 5.203 | 5.113 | 5.409 |
| 4 | N | 0 | 4-CH$_3$ | 3-Cl | 3.07 | 5.513 | 5.524 | 5.397 |
| 5 | N | 0 | 3-CH$_3$ | 3-Cl | 9.43 | 5.025 | 5.047 | 5.216 |
| 6 | N | 0 | 3-CH$_3$ | 4-NO$_2$ | 15.47 | 4.811 | 4.814 | 4.672 |
| 7 | N | 0 | 3,4-Me$_2$ | 3-Cl | 0.99 | 6.004 | 5.993 | 5.847 |
| 8 | N | 0 | 3,4-Me$_2$ | 3-CF$_3$ | 1.85 | 5.733 | 5.714 | 5.832 |
| 9 | N | 0 | 2-F | 3-Cl | 13.87 | 4.858 | 4.988 | 5.038 |
| 10* | N | 0 | 4-F | 3-Cl | 9.74 | 5.011 | 5.12 | 5.052 |
| 11 | N | 0 | 3-Cl | 3-Cl | 6.73 | 5.172 | 5.014 | 5.045 |
| 12* | N | 0 | 3,4-Cl$_2$ | 3-Cl | 6.05 | 5.218 | 5.078 | 4.98 |
| 13 | N | 0 | 3,4-Cl$_2$ | H | 17.62 | 4.754 | 4.805 | 4.759 |
| 14 | N | 1 | H | H | 31.5 | 4.502 | 4.707 | 4.73 |
| 15* | C | 1 | 3-Cl | H | 7.74 | 5.172 | 5.351 | 5.252 |
| 16 | C | 1 | 2-F | H | 14.11 | 4.85 | 4.837 | 4.833 |
| 17 | C | 1 | 4-CH$_3$ | H | 5.16 | 5.287 | 5.29 | 5.09 |
| 18 | C | 1 | 3-CH$_3$ | H | 7.39 | 5.131 | 5.16 | 5.202 |
| 19 |  | | | | 0.40 | 6.398 | 6.573 | 6.767 |
| 20 |  | | | | 0.09 | 7.046 | 6.896 | 6.674 |

**Table 1** (continued)
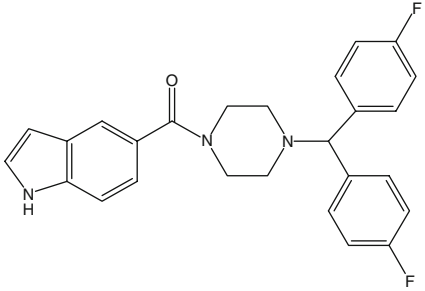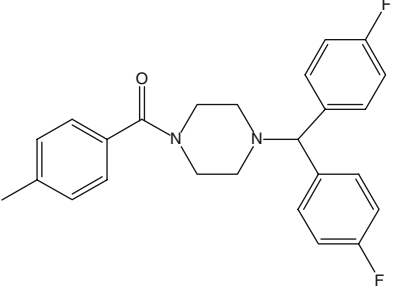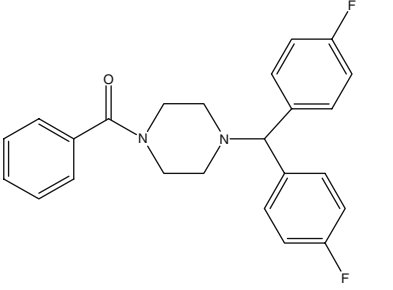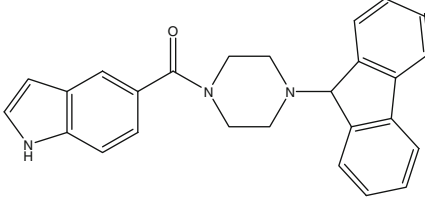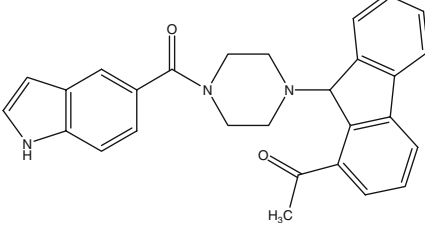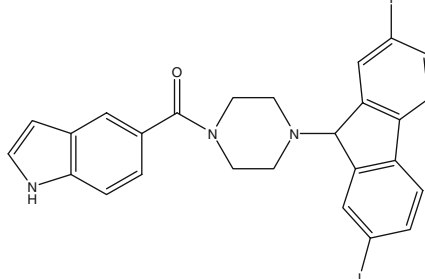
| | | | | | |
|---|---|---|---|---|---|
| 21 |  | 1.04 | 5.983 | 5.937 | 5.939 |
| 22* |  | 1.89 | 5.724 | 6.138 | 6.145 |
| 23 |  | 2.04 | 5.69 | 5.677 | 5.681 |
| 24 |  | 0.16 | 6.796 | 6.864 | 6.561 |
| 25 |  | 0.34 | 6.469 | 6.327 | 6.473 |
| 26 |  | 0.13 | 6.886 | 6.855 | 6.877 |

**Table 1** (continued)
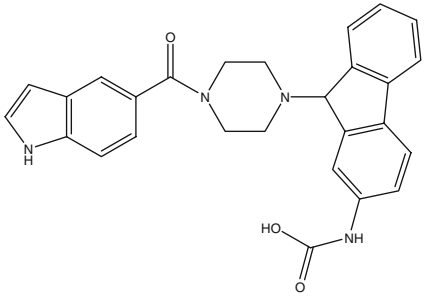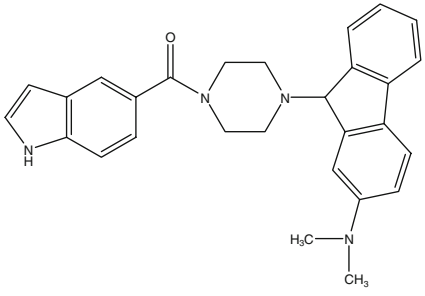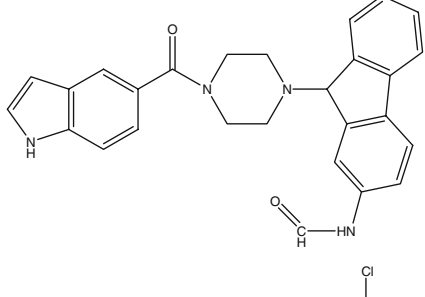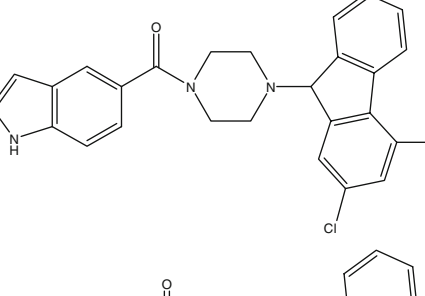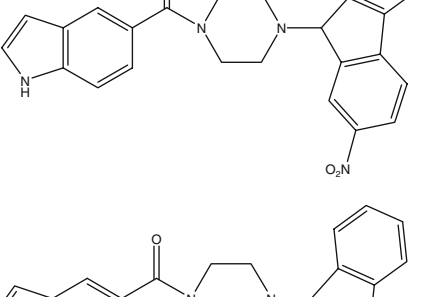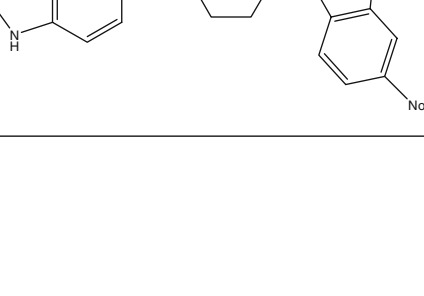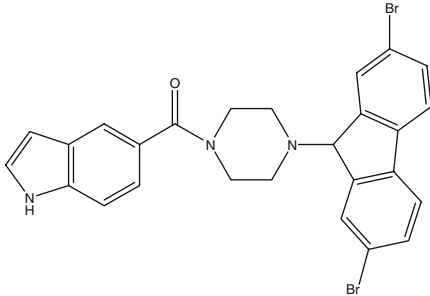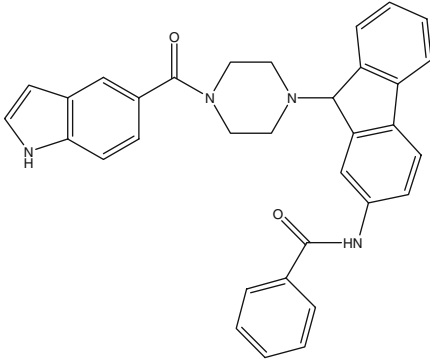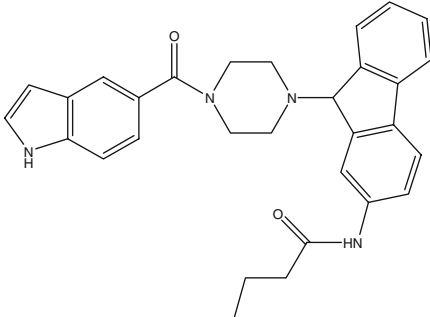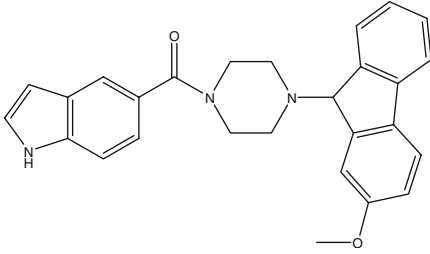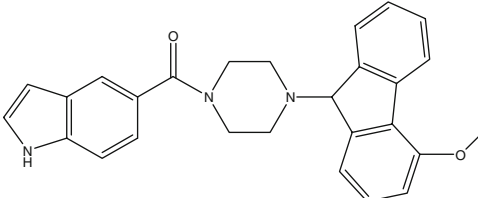
| | | | | | |
|---|---|---|---|---|---|
| 27 |  | 0.28 | 6.553 | 6.566 | 6.433 |
| 28 |  | 0.91 | 6.041 | 6.029 | 6.031 |
| 29 |  | 0.18 | 6.745 | 6.753 | 6.667 |
| 30* |  | 0.17 | 6.770 | 6.425 | 6.537 |
| 31 |  | 0.13 | 6.886 | 6.942 | 7.086 |
| 32 |  | 0.13 | 6.886 | 6.848 | 6.974 |

**Table 1** (continued)

| | | | | | |
|---|---|---|---|---|---|
| 33* |  | 0.12 | 6.921 | 6.783 | 6.586 |
| 34 |  | 0.59 | 6.229 | 6.168 | 6.269 |
| 35* |  | 0.46 | 6.337 | 6.573 | 6.479 |
| 36 |  | 0.52 | 6.284 | 6.433 | 6.309 |
| 37 |  | 0.52 | 6.086 | 6.005 | 5.871 |

was reflected by the third power of the atomic radii of the atoms. Electrostatic properties were introduced as atomic charges resulted from molecular docking. An atom-based hydrophobicity was assigned according to the parameterization developed by Ghose *et al*. [16]. The lattice dimensions were selected with a sufficiently large margin (>4Å) to enclose all the binding conformations of the inhibitors. In general, similarity indices, $A_{F,K}$ between the compounds of interest were computed by placing a probe atom at the intersections of the lattice points using Eq. 1.

$$A_{F,K(j)}^{q} = -\sum_{i=1}^{n} W_{\mathrm{probe},k}\ W_{ik}\mathrm{e}^{-a}r_{iq}^{2} \qquad (1)$$

Where $q$ represents a grid point, $i$ is the summation index over all atoms of the molecule $j$ under computation, $W_{ik}$ is the actual value of the physicochemical property $k$ of atom $i$, and $W_{\mathrm{probe},k}$ is the value of the probe atom. In the present study, similarity indices were computed using a probe atom ($W_{\mathrm{probe},k}$) with charge +1, radius 1Å, hydrophobicity +1 and attenuation factor a of 0.3 for the Gaussian type distance. The statistical evaluation for the CoMSIA analyses was performed in the same way as described for CoMFA.

### 3D Pharmacophore search

A 3D pharmacophore query with partial match constraints was defined on the basis of the structural features of the MtENR inhibitor Genz-10850 by using the crystal structure bound conformation (PDB entry code 1P44). The 3D pharmacophore search was performed by using the unity flexible search protocol as implemented in Sybyl7.1, with all options set as default [17]. In the unity search, the conformations of the screening database were generated on the fly by means of the Directed tweak method [18].

### Results and discussion

### CoMFA and CoMSIA analysis

Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) are 3D-QSAR methods intended to correlate the molecular features of a series of compounds with biological activities. CoMFA and CoMSIA do not take into account receptor ligand interactions but rely only on the calculation of molecular fields of the ligands and their subsequent correlation, by PLS regression, to their biological activities. In particular, CoMFA is based on the calculation of steric and electrostatic fields, while CoMSIA also considers hydrophobic, hydrogen bond acceptor and hydrogen bond donor fields. CoMFA and CoMSIA may be extremely sensitive to molecular alignment rules and overall consistency of the molecular alignment, therefore, the determination of spatial molecular alignment is a crucial step in 3D-QSAR studies, since the analyses are highly dependent on the quality of the alignments [19, 20]. Thus molecular docking studies were carried out using the program Flexx in order to generate receptor based alignment for model building. Combining molecular docking with 3D-QSAR modeling offers a more interesting, integrated approach and allows us to utilize structural information of the protein for 3D-QSAR modeling. The advantage of a docking-based model is that we can directly superimpose the contour plots into the protein active site. Such superimposition will also allow us to check the correlation between the contour plots and the corresponding receptor residues present near them. Because the docked pose gives the bioactive conformation of the ligands, this method helps to overcome the error that may arise by using an incorrect conformation of the ligand. We used X-ray structure of enzyme from the MtENR-Genz10850 complex (PDB code-1P44) [5] for molecular docking studies. The hydrophobic binding pocket of MtENR is made up of key residues Gly96, Phe97, Ile102, Met103, Phe149, Met155, Pro156, Ala157, Tyr158, Met161, Pro193, Met199, Val203, Ile215, Leu218 and Trp222. To ensure that the ligand orientation and the position obtained from the docking studies were likely to represent valid and reasonable binding modes of the inhibitors and virtual screening hits, the FlexX program docking parameters were first validated for the crystal structure used (1P44). The results of docking showed that FlexX successfully reproduced the crystal structure conformation of Genz-10850 with rmsd of 0.54Å. Moreover, the correlation between the FlexX scores and biological activities of arylamides was also good. Highly active compounds were found to be among the top ranking compounds according to the FlexX scores. Molecular docking of arylamides into *Mt*ENR binding site revealed very clear preference for the inhibitor binding pocket and all the compounds occupy the same spatial position as the co-crystallized *Mt*ENR inhibitor Genz10850. Almost all of the inhibitors bind in similar fashion with its substituted aromatic ring alongside piperizine and piperidine scaffold occupying the interior of the deep cleft and aromatic substitution toward urea is extended to the entrance of the hydrophobic binding cavity as shown in Fig. 1. The formation of two strong hydrogen bonds between ligand oxygen atom and active site residue Tyr158 and NADH cofactor probably serves as the key feature that governs the orientation of a compound within the active site. This hydrogen bonding network seems to be a conserved feature among all the InhA inhibitor complexes identified so far [7]. The same hydrogen bonding network was observed in
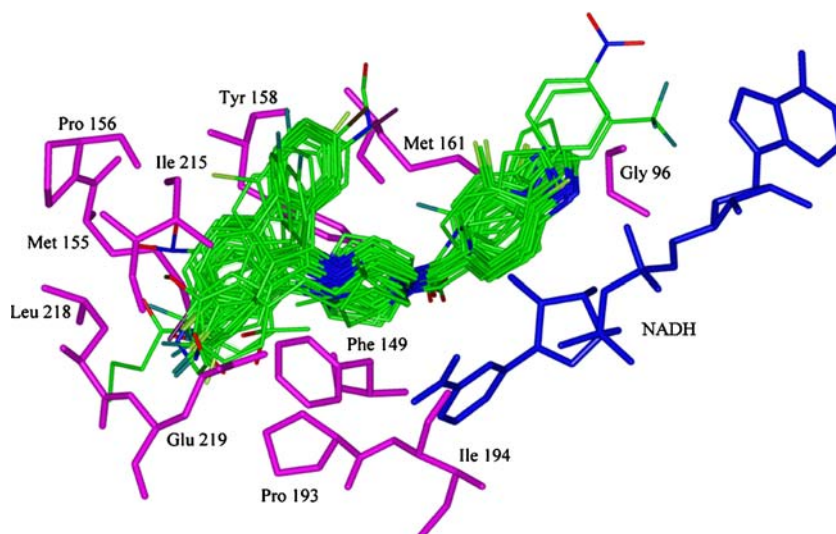
the crystal structure complexes of Genz-10850, triclosan and pyrrolidine carboxamides with InhA [5, 7]. In addition with hydrogen bonding interactions, π-π interaction between nicotinamide ring of cofactor NADH and at least one of the rings in the ligand contribute significantly toward inhibitor binding to InhA active site. This π-π interaction also seems to be conserved among the ENR's of various species. Some hydrophobic interactions with Phe149, Pro193, Leu218 and Val203 are also important with respect to inhibitor binding. All of the structural insights obtained from molecular docking of aryl amides are consistent with the available experimental activity data [5–8], suggesting that the enzyme-inhibitor binding structures obtained from the molecular docking are reasonable and can be used for developing CoMFA and CoMSIA models.

Twenty nine of 37 MtENR inhibitors were randomly picked up as training set for constructing the CoMFA and CoMSIA models. The remaining eight inhibitors were used as test set for model validation. Partial least square analysis was carried out for the training set molecules and the results are presented in Table 2 which shows that CoMFA model with LOO cross-validated $q^2$ value of 0.663 and CoMSIA model with $q^2$ value of 0.639 using five components was obtained. The non cross-validated PLS analysis resulted in conventional $r^2$ of 0.989, F value of 405.981 and standard error value of 0.098 for CoMFA model and $r^2$ of 0.963, F value of 120.871 and standard error value of 0.176 for CoMSIA model. In CoMFA, the steric field descriptors explain 60.9% of the variance while electrostatic field descriptors explain 39.1% of the variance whereas in case of CoMSIA, the steric field descriptors explain 20.2%, electrostatic field descriptors explain 32.3%, hydrophobic field descriptors explain 27.8% and hydrogen bond acceptor explain 19.7% of the variance. It has been established that five different descriptor fields are not totally indepen-

dent of each other and that such dependency among individual fields decreases the statistical significance of the models [21, 22]. An evaluation of which is actually needed for the generation of predictive model was performed by computing all possible combination of fields.

**Table 2** Statistical results for CoMFA and CoMSIA models

|                                  | CoMFA   | CoMSIA  |
|----------------------------------|---------|---------|
| **PLS Statistics**               |         |         |
| $q^{2a}$                         | 0.663   | 0.639   |
| $r^{2b}$                         | 0.989   | 0.963   |
| $S^c$                            | 0.098   | 0.176   |
| $F^d$                            | 405.981 | 120.871 |
| Optimal components$^e$           | 5       | 5       |
| Field distribution %             |         |         |
| Steric                           | 60.9    | 20.2    |
| Electrostatic                    | 39.1    | 32.3    |
| Hydrophobic                      | –       | 27.8    |
| Hydrogen bond acceptor           | –       | 19.7    |
| $r^2_{bootstrap}{}^f$            | 0.993   | 0.980   |
| $S_{bootstrap}{}^g$              | 0.074   | 0.134   |
| $r^2_{pred}{}^h$                 | 0.882   | 0.875   |
| $S_{pred}{}^i$                   | 0.240   | 0.247   |

$^a$ Cross-validated correlation coefficient

$^b$ Non-cross-validated correlation coefficient

$^c$ Standard error of estimate

$^d$ F test value

$^e$ Optimum number of components

$^f$ Bootstrap correlation coefficient

$^g$ Standard error of Bootstrapping

$^h$ Predictive correlation coefficient

$^i$ Standard error of prediction

The model with steric, electrostatic, hydrophobic and hydrogen bond acceptor field appeared to be superior among all the models derived thus final CoMSIA model was derived using these four field combinations. To further access the robustness and statistical significance of the derived models, bootstrap analysis for 100 runs was then carried out for further validation of the model by statistical sampling of the original data set to create new data sets. Thus, the difference in the parameters calculated from the original data and the average of the parameters calculated from the N (=100) runs of bootstrapping sampling is a measure of the bias of the original calculation. This yielded higher $r^2_{bootstrap}$ value 0.993 and 0.980 with standard error value of 0.074 and 0.134 for CoMFA and CoMSIA respectively and further supports the statistical validity of the developed models. The predicted activities for the inhibitors versus their experimental activities are listed in Table 1 and the correlation between the predicted activities and the experimental activities is depicted in Fig. 2a and b. Table 2 and Fig. 2a and b demonstrate that the predicted activities by the constructed CoMFA and CoMSIA models are in good agreement with the experimental data, suggesting that the CoMFA and CoMSIA models should have a satisfactory predictive ability.

The eight randomly selected compounds (compounds marked as * in Table 1) were used as test set to verify the stability and predictive ability of the constructed CoMFA and CoMSIA models. The predicted IC$_{50}$ (pIC$_{50}$) are in good agreement with the experimental data within a statistically tolerable error range, with a predictive correlation coefficient of $r^2_{pred}$=0.882 and 0.240 as the standard error of predictions for CoMFA and $r^2_{pred}$ = 0.875 and 0.247 as the standard error of predictions for CoMSIA models (Table 2). The correlation between the CoMFA and CoMSIA predicted activities and the experimental activities of the test set compounds are depicted in Fig. 2a and b. The test set results indicate that the CoMFA and CoMSIA models would be reliably used in predicting the activity of new compounds.

CoMFA and CoMSIA contour analysis

Since the QSAR models were built on the basis of docking, we could overlay the 3D molecular fields or contour maps produced by CoMFA and CoMSIA into the receptor binding pocket, thus evaluating the complementarity of the ligand based QSAR and our proposed model for ligand receptor interactions. The contour maps were generated as scalar products of coefficients and standard deviation, associated with each CoMFA or CoMSIA column. The maps generated depict regions having scaled coefficients greater than 80% (favored) or less than 20% (disfavored). In the case of CoMFA, the green contour shows favorable
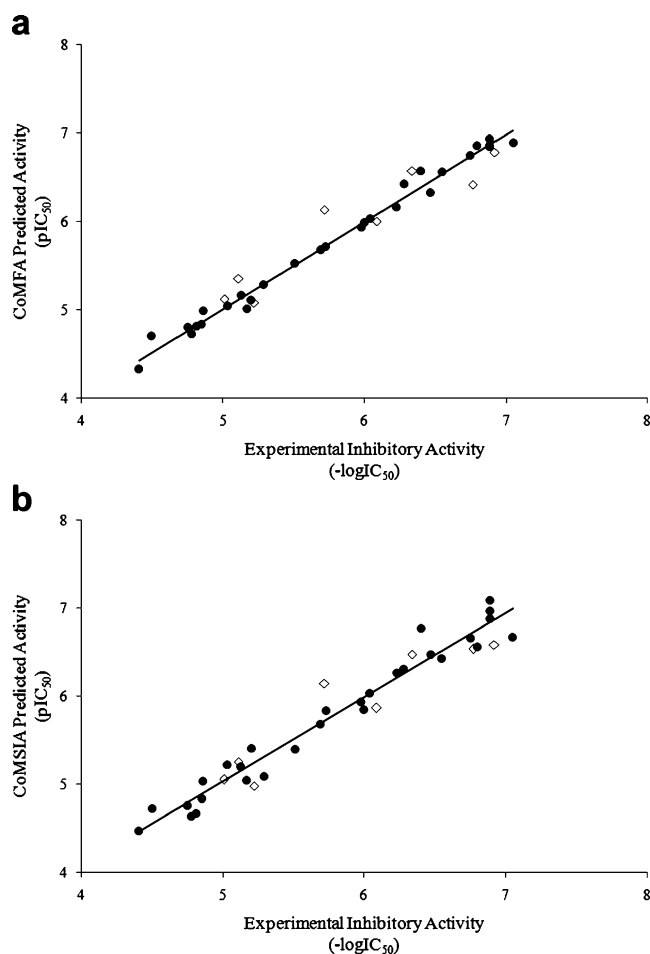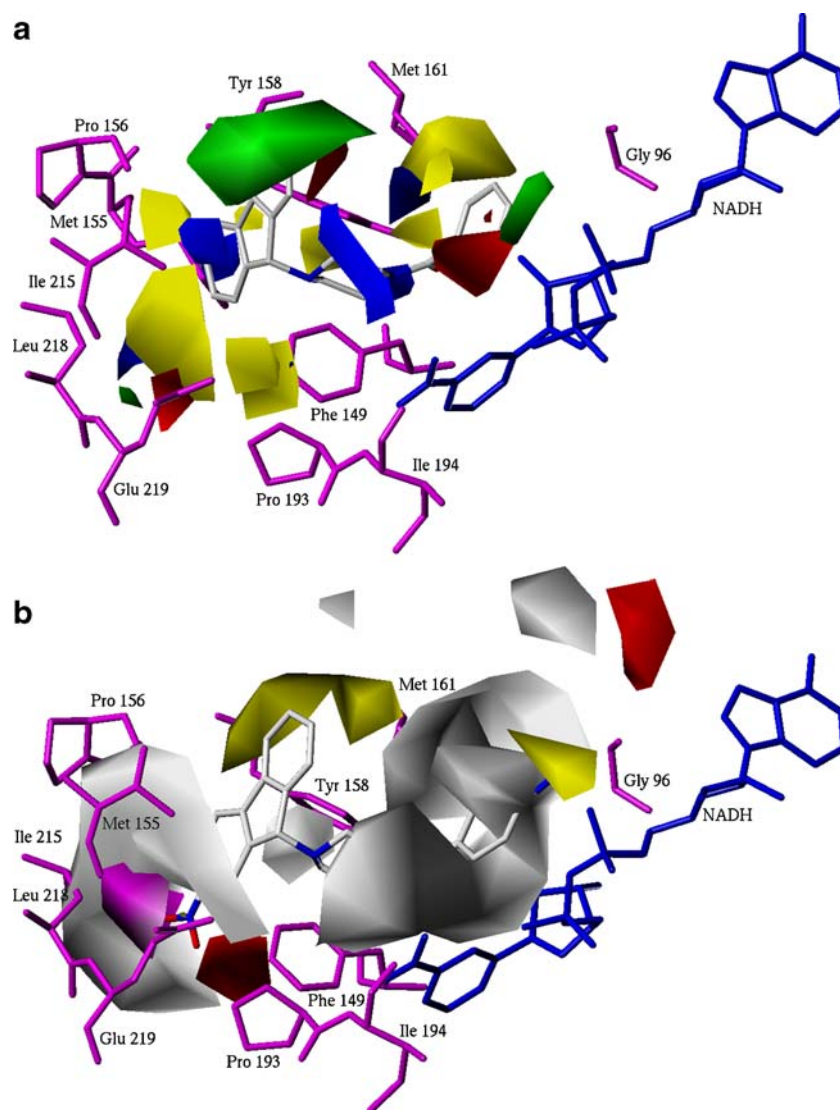


Fig. 2 Correlation between predicted activities and experimental activities of training and test set compounds (a) CoMFA (b) CoMSIA

steric interaction and the yellow contours show the region where the steric group is not favored. The red contours show favorable electronegative regions, and the blue contour shows the region where the electropositive region is favored. One of the most active compounds in the series (compound 20) is shown superimposed with the CoMFA steric and electrostatic contour maps in Fig. 3a. The green contour region near the polyaromatic fluorene ring of compound 20 shows that bulky substituents at this position have favorable steric interactions. This is consistent with the reported X-ray crystal structure of MtENR bound with Genz-10850 (PDB ID: 1P44) where fluorene ring is sandwiched between Phe149, Pro156, Tyr158 and Ile215 and that is why compounds 19, 24–37 having the substituted and unsubstituted fluorene ring at similar position have higher activity. The removal of bulky groups at this position like in the case of compound 1–6 and compound 9–18 lead to decreased activity when compared with compounds with bulky poly aromatic ring. The yellow contours are present below the plane of the fluorene ring so there would be an unfavorable steric interaction if sub-
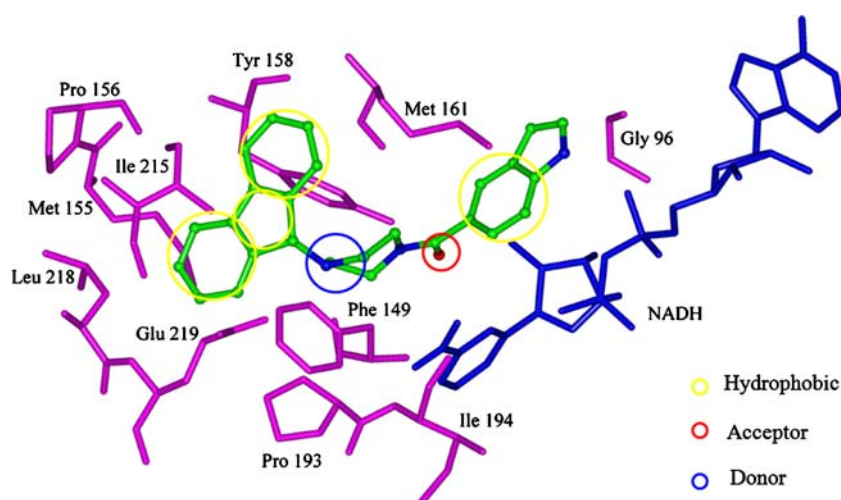
stituents protrude in that region. This is reason for compound **14** and compounds **15–18** with an extra methyl group attached to piperizine and piperidine ring showing lower inhibitory activity. A blue contour region (favorable electropositive group) is found in overlapping piperizine nitrogen in compound **20** as shown in Fig. 3a which shows electropositive character is favorable for inhibitory activity at this position. The substitution of piperidine ring in place of piperizine in compounds **15–18** leads to decrease in activity of these compounds. Similarly, a red contour region (favorable electronegative group) overlapping amide carbonyl states the reason for inhibitory activity of amides. The amide oxygen atom here is involved in conserved hydrogen bonding interaction with side chain oxygen of Tyr158 and 2′ oxygen of ribose moiety of cofactor NADH.

In case of CoMSIA, we get additional insight from the hydrophobic and hydrogen bond acceptor features. The

CoMSIA steric and electrostatic contour maps were similar to the ones obtained from the CoMFA model. Fig. 3b shows compound **31** superimposed on the CoMSIA hydrophobic and hydrogen bond acceptor contour plots. In Fig. 3b, a white contour showing an unfavorable hydrophobic interaction region near amino acid residues Ile215, Leu218 and Glu219. If any substituent extends to this region it will have close hydrophobic clashes with Ile215 and Leu218. This explains the lower activity of compounds **1–6** and **9–18** compared to that of compounds **19**, **20**, **24–37** where hydrophobic substituents protude toward favorable hydrophobic interaction yellow region. The hydrogen-bond acceptor contour superimposed on compound **31** is shown in Fig. 3b. The magenta contour shows a favorable hydrogen-bond acceptor region, and the red contour shows regions in which the hydrogen-bond acceptor is not favored. From the contour plot, we can see that nitro group substituted at 2 and 3 position of polyaromatic fluorene ring in compounds **31**

and **32** is present in the favorable hydrogen-bond acceptor group region which explains the high activity of these compounds.

### 3D-Pharmacophore search and molecular docking

We adopted 3D crystal structure of MtENR in complex with potent inhibitor Genz-10850 (PDB entry code 1P44) to generate 3D pharmacophore hypothesis. We defined the 3D pharmacophore on the basis of crystal structure bound conformation of Genz-10850 using partial match constraints and grouping the pharmacophoric features into two clusters. In the pharmacophore search, the screened compounds were required to match at least two hydrophobic features from first cluster to be considered hits (Fig. 4). The first cluster included three hydrophobic features, which in our hypothesis form hydrophobic interactions with Phe 149, Pro193, Ile 215 and Leu 218. The second cluster includes a hydrogen bond donor, a hydrogen bond acceptor and a hydrophobic feature and hits were required to match the entire features from this cluster. Hydrogen bond acceptor is very important for MtENR inhibition as it is involved in hydrogen bonding interaction with side chain OH of Tyr 158 and 2′ oxygen of ribose moiety of NADH. This particular interaction is very important and has been conserved in most ENR's for which the crystal structure has been solved in complex with NADH. We have considered a hydrogen bond donor in our pharmacophore hypothesis, though while considering the crystal structure bound ligand and other piperazine based compounds there was no need to include this pharmacophore feature as it does not have nearby interaction partners in the binding site. The nearest acceptor atoms are side chain oxygen of Glu219 and Tyr 158, backbone oxygen of Ile 194 and carbonyl group of cofactor NADH which are 5.3, 6.2, 6.3 and 4.6Å away, respectively. We have included a donor here in the pharmacophore hypothesis for two reasons, the first

reason is to look for piperizine based compounds which actually possess MtENR activities and presence of nitrogen rich environment is necessary for the activity and the second and major reason is to look for a
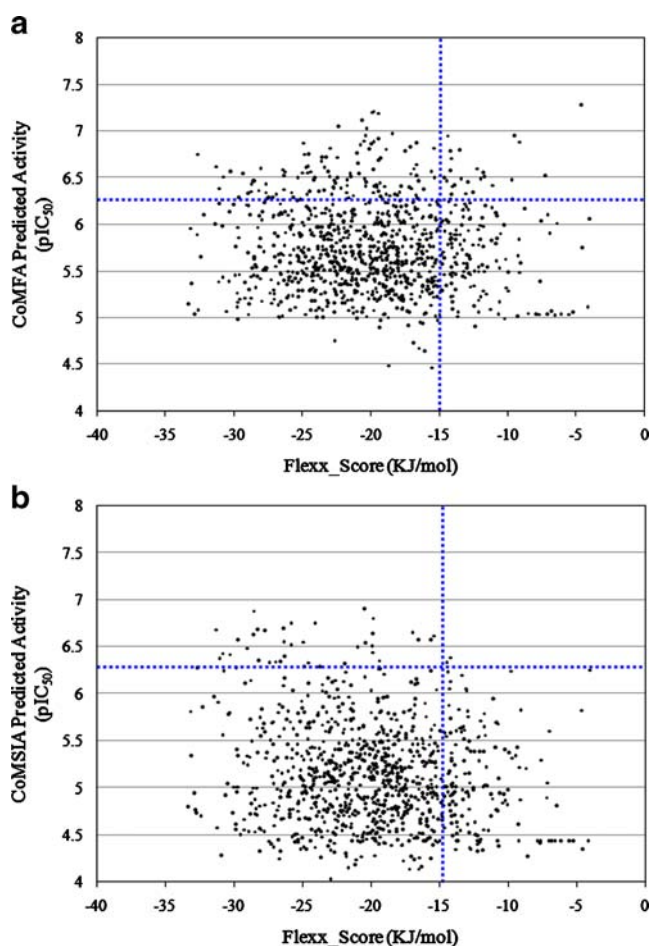


Fig. 5 Workflow of CoMFA and CoMSIA based compound selection. A $pIC_{50}$ of 6.25 and Flexx_Score of -15.00 kJ mol$^{-1}$ was taken as cutoff for the selection of compounds. (a) CoMFA based selection (b) CoMSIA based selection

**Table 3** Chemical structures of putative binders and their respective docking scores, CoMFA and CoMSIA predicted activity, molecular weight and LogP identified using virtual screening and QSAR
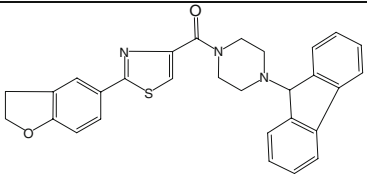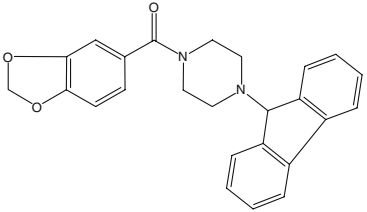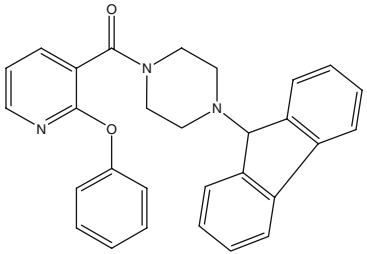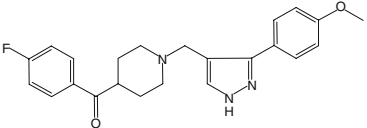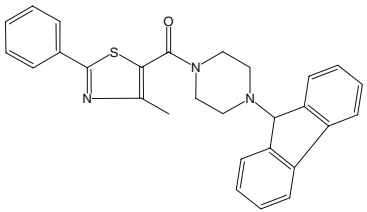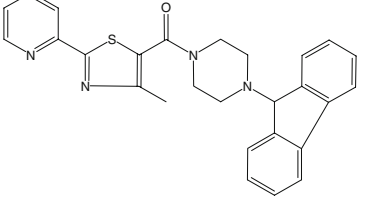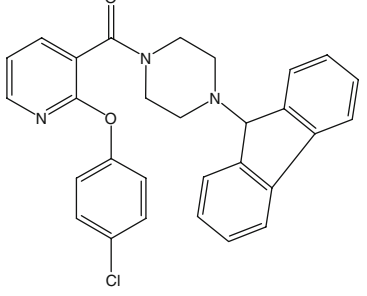
| S. No | Hits | Structure | Flexx_Score (kJ/mol) | CoMFA pIC50 | CoMSIA pIC50 | Mol. Wt | LogP (o/w) |
|-------|------|-----------|----------------------|-------------|--------------|---------|------------|
| 1 | HTS08262 | | -31.38 | 6.62 | 6.685 | 479.60 | 5.08 |
| 2 | HTS07943 | | -31.07 | 6.392 | 6.374 | 398.46 | 4.32 |
| 3 | HTS07936 | | -30.83 | 6.269 | 6.432 | 447.54 | 5.39 |
| 4 | HTS05698 | | -28.98 | 6.332 | 6.486 | 393.46 | 4.40 |
| 5 | HTS07939 | | -28.57 | 6.462 | 6.885 | 451.59 | 5.36 |
| 6 | HTS09780 | | -28.53 | 6.463 | 6.622 | 453.57 | 3.24 |
| 7 | HTS07938 | | -27.44 | 6.289 | 6.324 | 481.98 | 5.99 |

**Table 3** (continued)
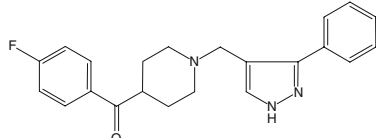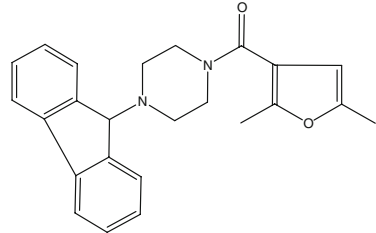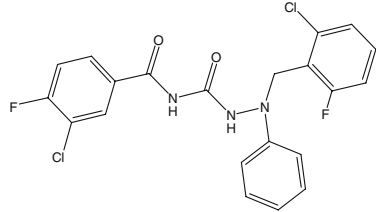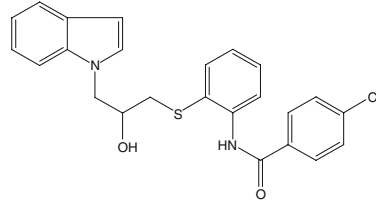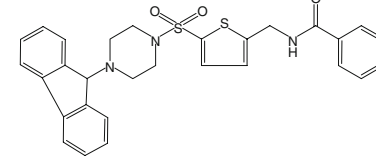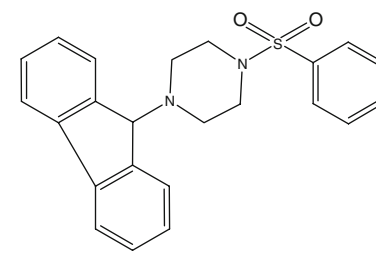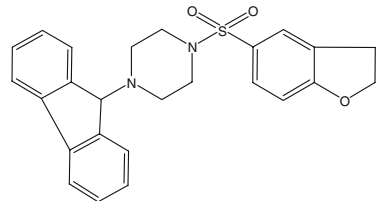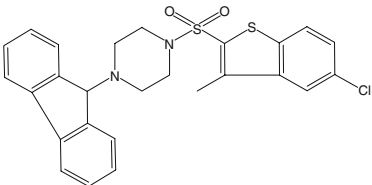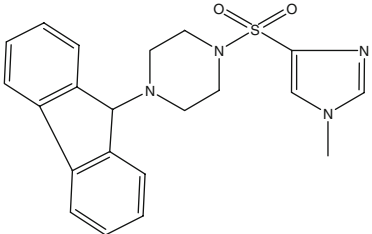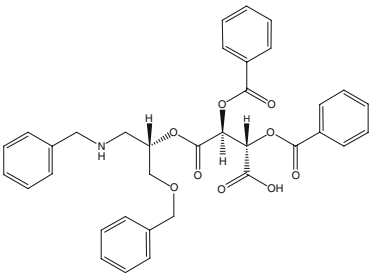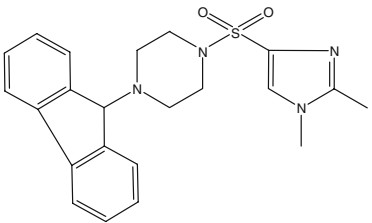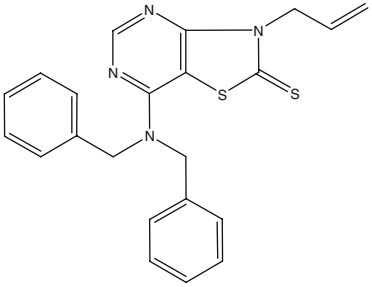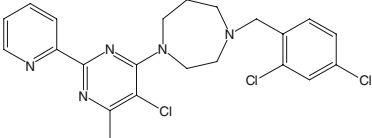
| 8 | HTS05682 |  | -26.97 | 6.316 | 6.342 | 363.44 | 4.44 |
| 9 | HTS07942 |  | -26.38 | 6.587 | 6.549 | 372.47 | 3.84 |
| 10 | BTB06014 |  | -26.34 | 6.516 | 6.689 | 450.27 | 5.34 |
| 11 | HAN00324 |  | -26.27 | 6.566 | 6.392 | 436.96 | 5.84 |
| 12 | HTS08278 |  | -24.88 | 6.875 | 6.554 | 529.69 | 4.69 |
| 13 | HTS08766 |  | -21.85 | 6.281 | 6.314 | 390.51 | 3.86 |
| 14 | HTS08272 |  | -20.99 | 6.757 | 6.49 | 432.54 | 3.76 |

**Table 3** (continued)

| 15 | HTS08762 |  | -20.43 | 6.943 | 6.898 | 495.07 | 6.08 |
| 16 | HTS08274 |  | -19.93 | 6.86 | 6.502 | 394.50 | 2.39 |
| 17 | BTB15159 |  | -19.86 | 7.206 | 6.808 | 611.65 | 5.40 |
| 18 | HTS08761 |  | -19.82 | 6.88 | 6.633 | 408.53 | 2.72 |
| 19 | HTS09058 |  | -16.48 | 6.464 | 6.566 | 404.56 | 4.99 |
| 20 | AW01237 |  | -15.35 | 6.818 | 6.618 | 462.81 | 4.79 |

compound which may form additional hydrogen bonding interactions with Glu219, Ile194 and NADH as there exists possibilities of hydrogen bonding with these residues. The pharmacophore hypothesis was then used as a 3D structural query to retrieve molecules matching 3D query from Maybridge small molecule database. Virtual screening was carried out using flex utility of UNITY module available with Sybyl7.1. The screening of the pharmacophore query yielded 996 hits that met the specified requirements. All the hit compounds retrieved from the pharmacophore based screening were then subjected to molecular docking to the MtENR binding site in order to select the compounds on the basis of their ability to form favorable interactions with InhA active site. InhA-NADH-Genz10850 ternary complex (PDB entry code 1P44) was used for docking studies. The molecular docking was carried out using FlexX program and 30 distinct poses of each ligand in the active site were generated. As, FlexX fails to find docking solution of 22 out of 996 hits subjected to molecular docking, compound pose yielding the highest FlexX score for each 974 hits was selected for further analysis.

Selection of putative MtENR binders

Ligand based methods of analysis such as CoMFA and CoMSIA are widely used not only because they are not very computationally intensive but also they can lead to rapid generation of QSAR models from which the biological activity of new compounds can be predicted. In contrast, an accurate prediction of activity of untested compounds based on the computation of binding free energy is both complicated and lengthy. Overall, the CoMFA and CoMSIA results of the training and test set demonstrated good accuracy of the developed solutions,

their useful synergy, and ability to enrich for the most active target binders. These observations encouraged us to apply developed models to structure based virtual screening hits for the prediction of biological activity. Thus based on the findings derived from the developed CoMFA and ComSIA models, the biological activity of 974 hits was predicted with both CoMFA and CoMSIA models. As described in the previous section, CoMFA and CoMSIA models were based on Flexx docked poses. Therefore, in order to apply pretrained CoMFA and CoMSIA models best docked poses according to FlexX_Score were used. To further expand the utility of the developed QSAR models, we have implemented the consensus approach for selection of putative MtENR binders taking the top results identified separately by CoMFA, CoMSIA and Flexx scoring. The selection of compounds is illustrated in Fig. 5. As FlexX tries to determine the binding free energy, hits which have a good FlexX score, $i.e.$, more than -15 kJ mol$^{-1}$ were selected. As the moderate to highly active MtENR inhibitors fall in –logIC$_{50}$ of ~6.25 to ~7.0 so CoMFA and CoMSIA predicted activity (pIC$_{50}$) of 6.25 was taken as cutoff for the selection of compounds. Thirty four compounds meeting the above mentioned criteria were then visually inspected to incorporate additional experimental knowledge, $i.e.$, (1) $\pi$ stacking interactions with the nicotinamide ring of the cofactor NADH, (2) hydrogen bonding with 2′ hydroxyl moiety of the ribose and hydroxyl group of active site residue Tyr 158 and (3) hydrophobic interactions with Phe149, Pro193, Leu218 and Val203. After the final visual inspection, we formed a list of 20 compounds out of which seven compounds belong to the arylamides class of chemical compounds which are known to possess MtENR inhibitory activity. Thus, our virtual screening study outlines novel point of substitution for rational design of arylamides for further improving the biological activity. The remaining 13 compounds could not



**Fig. 6** Binding mode of HTS08762 (atom color) shown with Genz10850 (brown). Dashed lines indicate hydrogen bond between HTS08762 and Tyr158 and NADH
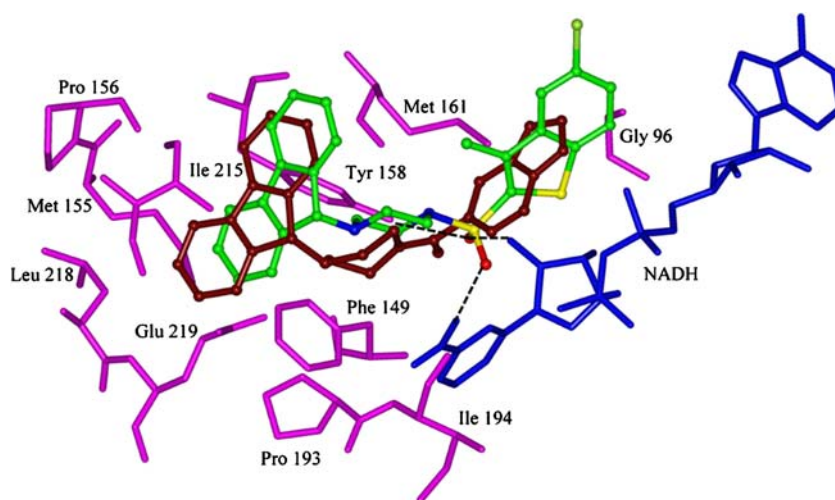
be found in the scientific literature as ever tested against MtENR. The structures of the selected compounds are shown in Table 3, together with their corresponding Flexx_Score, CoMFA and CoMSIA predicted activity, molecular weight and calculated log P values. The promising compounds share common pharmacophoric features like a nitrogen-rich scaffold (sulfonyl piperizine, pyrazolopiperidine, amides, benzylamides, diazepane, benzyl ester *etc*.) and hydrophobic group connected to the scaffold. Among these sulfonyl piperizine scaffold forms the major class and could be exploited for further optimization. The predicted binding mode of a sulfonyl piperizine compound **HTS08762** displaying high CoMFA and CoMSIA predicted activity is described in Fig. 6 along with the crystal structure bound conformation of inhibitor Genz-10850 with MtENR. Like the bound conformation of Genz-10850 shown in brown color in Fig. 6, **HTS08762** is anchored to the cavity by combination of π stacking interactions with the nicotinamide ring of the cofactor NADH and hydrogen bonding interaction between sulfonyl oxygen and sidechain oxygen of Tyr158 and 2′ hydroxyl of the ribose. Along with these, hydrophobic interactions of ligand with Phe149, Pro193, Leu218 and Val203 also act as stabilizing force for protein ligand binding.

## Conclusions

Using available information on 37 known arylamides as MtENR inhibitors, we have developed receptor based CoMFA and CoMSIA structure–activity models using docked poses for molecular alignment. The contour maps from receptor-based both CoMFA and CoMSIA model shows that docking-based models generally match well with the active site of MtENR. The resulting CoMFA and CoMSIA models combined with structure-based virtual screening allowed identification of several scaffolds like sulfonyl piperizine, pyrazole piperidine, diazepane *etc*. that have not been previously characterized in the scientific literature as MtENR inhibitors. Since MtENR represents a prospective drug target against tuberculosis, the identified putative MtENR binders can be characterized as potential therapeutic agents laying a foundation for future lead identification and optimization studies.

## References

1. Takayama K, Wang C, Besra GS (2005) Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis. Clin Microbiol Rev 18:81–101
2. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, Collins D, de Lisle G, Jacobs WR (1994) *inhA*, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science 263:227–230
3. Zhang Y, Heym B, Allen B, Young D, Cole S (1992) The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature 358:591–593
4. Escalante P, Ramaswamy S, Sanabria H, Soini H, Pan X, Valiente-Castillo O, Musser JM (1998) Genotypic characterization of drug-resistant *Mycobacterium tuberculosis* isolates from Peru. Tuber Lung Dis 79:111–118
5. Kuo MR, Morbidoni HR, Alland D, Sneddon SF, Gourlie BB, Staveski MM, Leonard M, Gregory JS, Janjigian AD, Yee C, Musser JM, Kreiswirth B, Iwamoto H, Perozzo R, Jacobs WR, Sacchettini JC, Fidock DA (2003) Targeting tuberculosis and malaria through inhibition of Enoyl reductase: compound activity and structural data. J Biol Chem 278:20851–20859
6. Sullivan TJ, Truglio JJ, Boyne ME, Novichenok P, Zhang X, Stratton CF, Li H, Kaur T, Amin A, Johnson F, Slayden RA, Kisker C, Tonge PJ (2006) High affinity InhA inhibitors with activity against drug-resistant strains of Mycobacterium tuberculosis. ACS Chem Biol 1:43–53
7. He X, Alian A, Stroud R, Ortiz de Montellano PR (2006) Pyrrolidine carboxamides as a novel class of inhibitors of enoyl acyl carrier protein reductase from Mycobacterium tuberculosis. J Med Chem 49:6308–6323
8. He X, Alian A, Ortiz de Montellano PR (2007) Inhibition of the Mycobacterium tuberculosis enoyl acyl carrier protein reductase InhA by arylamides. Bioorg Med Chem 15:6649–6658
9. Kumar A, Chaturvedi V, Bhatnagar S, Sinha S, Siddiqi MI (2009) Knowledge based identification of potent antitubercular compounds using structure based virtual screening and structure interaction fingerprints. J Chem Inf Model 49:35–42
10. Rarey M, Kramer B, Lengauer T, Klebe GA (1996) Fast flexible docking method using an incremental construction algorithm. J Mol Biol 261:470–489
11. Clark MC, Cramer RD III, van Opden Bosch N (1989) Validation of the General Purpose Tripose 5.2 Force Field. J Comput Chem 10:982–1012
12. Wold S, Ruhe A, Wold H, Dunn WJI (1984) The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses. SIAM J Sci Stat Comput 5:735–743
13. Wold S, Albano C, Dunn WJ III, Edlund U, Esbensen K, Geladi P, Hellberg S, Johanson E, Lindberg W, Sjostrom M (1984) Multivariate data analysis in chemistry. NATO ASI Ser Ser C 138:17–95
14. Clark M, Cramer RD III (1993) The Probability of Chance Correlation Using Partial Least Squares (PLS). Quant Struct-Act Relat 12:137–145
15. Bush BL, Nachbar RB (1993) Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. J Comput-Aided Mol Des 7:587–619
16. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295:337–356
17. SYBYL Molecular Modeling System, Version 7.1 (2005) Tripos Inc, St Louis, MO
18. Hurst T (1994) Flexible 3D searching - the directed tweak technique. J Chem Inf Comput Sci 34:190–196

19. Salum LB, Polikarpov I, Andricopulo AD (2007) Structural and Chemical Basis for Enhanced Affinity and Potency for a Large Series of Estrogen Receptor Ligands: 2D and 3D QSAR Studies. J Mol Graphics Modell 26:434–442

20. Honorio KM, Garratt RC, Polikarpov I, Andricopulo AD (2007) 3D QSAR Comparative Molecular Field Analysis on Nonsteroidal Farnesoid X Receptor Activators. J Mol Graphics Modell 25:921–927

21. Bringmann G, Rummey C (2003) 3D QSAR Investigations on antimalarial naphthylisoquinoline alkaloids by comparative molecular similarity indices analysis (CoMSIA), based on different alignment approaches. J Chem Inf Comput Sci 43:304–316

22. Bohm M, Sturzebecher J, Klebe G (1999) 3D QSAR analyses using CoMFA and CoMSIA to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. J Med Chem 42:458–477