3D-QSAR studies of triazolopyrimidine derivatives of *Plasmodium falciparum* dihydroorotate dehydrogenase inhibitors using a combination of molecular dynamics, docking, and genetic algorithm-based methods

Priyanka Shah · Sumit Kumar · Sunita Tiwari ·
Mohammad Imran Siddiqi

Received: 16 September 2011 / Accepted: 18 January 2012 / Published online: 5 February 2012
© Springer-Verlag 2012

Abstract A series of 35 triazolopyrimidine analogues reported as *Plasmodium falciparum* dihydroorotate dehydrogenase (*Pf*DHODH) inhibitors were optimized using quantum mechanics methods, and their binding conformations were studied by docking and 3D quantitative structure–activity relationship studies. Genetic algorithm-based criteria was adopted for selection of training and test sets while maintaining structural diversity of training and test sets, which is also very crucial for model development and validation. Both the comparative molecular field analyses ($q^2_{\text{LOO}} = 0.841$, $r^2_{\text{ncv}} = 0.99$) and comparative molecular similarity indices analyses ($q^2_{\text{LOO}} = 0.757$, $r^2_{\text{ncv}} = 0.943$) show excellent correlation and high predictive power. Furthermore, molecular dynamics simulations were performed to explore the binding mode of the two of the most active compounds of the series, **10** and **14**. Harmonization in the two simulation results validate the analysis and therefore applicability of docking parameters based on crystallographic conformation of compound **14** bound to receptor molecule. This work provides useful information

about the inhibition mechanism of this class of molecules and will assist in the design of more potent inhibitors of *Pf*DHODH.

Keywords QSAR · Genetic algorithm-based feature selection, hierarchical clustering, and docking · Molecular dynamics

Introduction

Quantitative structure–activity relationship (QSAR) studies seeks to construct a reliable model for the prediction of new data by exploring the relationship between molecular structures and experimental data. Robustness of QSAR model strongly relies on the qualities of the chemical structure information as well as statistical parameters used to produce relationship between structure and activity of molecules. The growing interest in QSAR and its unexplored enormous potential propelled development of newer statistical approaches and more suitable and novel physicochemical descriptors. In this evolving scenario, the good performance of 3D-QSAR methods, in particular, comparative molecular field analyses (CoMFA) and comparative molecular similarity indices analyses (CoMSIA) offered medicinal chemists a useful chance to visually appreciate the variation of molecular interaction fields, assessed by numerical chemical probes, and to fulfill the requirement and desire to predict specific biological responses [1].

However, the main limitations of ligand-based 3D-QSAR method have been the robustness and reliability of models being strongly dependent on the adopted criteria for conformation generation and molecular overlay. Ligand-based

P. Shah · M. I. Siddiqi (✉)
Molecular and Structural Biology Division,
Central Drug Research Institute,
Lucknow 226 001, India
e-mail: imsiddiqi@yahoo.com

S. Kumar
National Institute of Pharmaceutical Education and Research,
Raebareli, UP, India

S. Tiwari
Department of Physiology, C.S.M.M.U.,
Lucknow, UP, India

CoMFA method makes use of probe atoms, such as nitrogen, carbon, etc. to determine possible interaction fields between ligands and a putative receptor, and therefore, explicit atomic-level representation of receptor is not a prerequisite for these methods. Not considering the biological counterpart and, more importantly, the significant interactions determining ligand binding has raised important issues on the reliability of molecular alignments for structure–activity relationship study [2]. The necessary requirement of having an aligned data set imposes a fairly significant limitation. Sometimes optimal fitting of rigid body does not provide good predictive models because of the significant range of structural diversity of the compounds, the considerable size of some analogs, and conformational adjustments between receptor and ligand necessary to accommodate different ligands in the active site. The 3D-QSAR models obtained using such alignments may lead to poor external validation. Docking methods use real interaction field between the ligands and receptor, thus requiring advance knowledge of the receptor structure. Therefore, in those cases, where crystallographic structure of receptor molecule is solved, key interactions responsible for ligand–receptor binding in the active site of receptor molecule can efficiently be characterized with the help of molecular docking by offering predictions of the bound conformation for the ligand and a scheme for energetically scoring the ligand–receptor interaction [3]. However, experimentally determined affinities depend on several other factors including important dynamic or entropic effects that are difficult to strictly represent in a general scoring function; therefore, the docking score may not always correlate well with experimental data even with accurate structure predictions.

The current study deals with ligand-based and receptor-guided QSAR technique to characterize the binding pattern of triazolopyrimidine analogues in the active site of *Plasmodium falciparum* dihydroorotate dehydrogenase (PfDHODH). Of the four malarial parasites, *P. falciparum* causes the most severe form of malaria and accounts for over one million deaths annually. Pyrimidine biosynthesis presents an excellent target for development of new chemotherapeutic agents against the malaria parasite. Unlike mammalian cells, which contain enzymes for both de novo biosynthesis and salvage of preformed pyrimidine bases and nucleosides, the parasite relies exclusively on de novo synthesis. Dihydroorotate dehydrogenase is fourth enzyme in the pyrimidine biosynthetic pathway [5]. Dihydroorotate dehydrogenase is a mitochondrially localized flavoenzyme, which catalyzes the rate-limiting step of the oxidation of dihydroorotate (DHO) to orotate in the presence of the co-factors flavin mononucleotide (FMN) and ubiquinone (CoQ) in de novo pyrimidine biosynthesis pathway and is therefore an attractive antimalarial chemotherapeutic target [4].

Ojha and co-workers [6] have recently reported a QSAR study involving triazolopyrimidine derivatives where they

have reported that steric volume and charge distribution has important effect on the activity of o- m- and p-substituent of the phenyl ring attached to triazolopyrimidine group of PfDHODH inhibitors. In this work, we describe 3D-QSAR (CoMFA and CoMSIA) study for triazolopyrimidine analogue inhibitors of dihydroorotate dehydrogenase in *P. falciparum* in order to compare the information obtained from three-dimensional arrangements of atoms in the molecules with classical QSAR and extract more information in terms of steric and electrostatic properties from 3D-QSAR methods. It will be useful to build a QSAR model to predict and optimize the properties and activities of new untested triazolopyrimidine analogues and determine key structural requirements for their enhanced activity.

Effective selection of training set compounds is an important part of the QSAR modeling process. It has been indicated that to achieve the optimal model, the selection of training and test sets should be based on some rational algorithms; otherwise, poor predictive ability of QSAR models may be obtained [7]. Therefore, it is also an important step to select the group of molecules that represent the most critical structural and physicochemical features associated with activity. The predictive accuracy and confidence of a QSAR model for different unknown chemicals varies according to how well the training set represents the unknown chemicals and how robust the model is in extrapolating beyond the chemistry space defined by training set. In the present study, an attempt was made to rationalize the division process, in which the division was performed using hierarchical cluster analysis so that points representing both training and test sets were distributed within the whole descriptor space occupied by the entire dataset, and each point of the test set was close to at least 1 point of the training set. This procedure ensures that chemical classes will be represented in both series of compounds. (i.e., training and test sets). Genetic algorithm is a widely used algorithm based on the biological evolution and natural selection principles for optimization problems. Earlier, Yuan and co-worker [8] have successfully employed genetic algorithm for the CoMFA modeling to solve the selection of the ligand conformations based optimization problem. Depending on the operators of genetic algorithm, several different training and test sets were built and re-evaluated repeatedly and model showing statistically best compromise between internal and external validity was chosen for further analysis.

Two molecular dynamics (MD) simulations were performed: one involving the crystallographic bound conformation of compound 14 with the receptor protein pfDHODH (pdb id: 3I68) and the other with docked conformation of compound 10 bound to pfDHODH. MD studies provided better insight into the energetic stability of given bound ligand configurations of these two most active compounds.

Materials and methods

Data set

Figure 1 displayed structure of one of the most active triazolopyrimidine analogue compound 14. Thirty-five such novel inhibitors of PfDHODH were taken from the literature [9, 10] with their biological activities in terms of IC_{50} values [IC_{50} values, i.e., the concentration (μM) of inhibitor that produces 50% inhibition of PfDHODH], accordingly the pIC_{50} ($-\log IC_{50}$) are reported in Table 1.

Geometry optimization

Three-dimensional structures of 35 ligands were constructed using the SYBYL7.1 [11] suite of programs running under Irix 6.5. Full geometry optimization were calculated using B3LYP/STO-3G approach implemented in the source code of the general ab initio quantum chemistry package GAMESS [12] to determine a plausible stable conformational structure for the ligands.

Enzyme preparation

The crystal structure of PfDHODH complexed with compound 14 and cofactor FMN from Brookhaven Protein Data Bank (PDB ID code 3I68) was used in the docking experiments. Crystallographic waters, which were not hydrogen bonded to the enzyme, were deleted, and the complex was energy minimized by a 500-step steepest descent method with GROMACS v.4.0.5 [13]. Energy minimizations were realized by setting a 10-Å non-bonded cutoff and a 0.01-kcal/mol energy gradient convergence criterion. So far, all these steps were done by using the Gromos force field [14]. Finally, the minimized complex was used as the starting structure in the docking study.

Fig. 1 Structure of the compound 14, used as a template for alignment based on atoms highlighted in black

Docking

The binding modes of several triazolopyrimidine derivatives into the active site of energy minimized dihydroorotate dehydrogenase receptor were investigated using flexible docking with FlexX [15] to orient and score small molecules for shape and chemical complementarity to a macromolecular binding site. FlexX considers ligand conformational flexibility by an incremental fragment placing technique. For each ligand, the pose for the further study was selected on the basis of having the highest ChemScore [16], with the further stipulation that the following knowledge-based criteria (as determined by visual inspection) must be obeyed whenever possible: (1) good π – π overlap with residue Phe227, as has been found to be critical for binding; (2) within hydrogen bond distance with residues His185 and Arg265 as have also been found to be very important for complexation.

Structure alignment

CoMFA results are extremely sensitive to the alignment rules, overall orientation of the aligned compounds, lattice shifting step size, probe atom type, etc. Thus, atom fit molecular alignment method was employed in the present study. This method involves atom based fitting [root mean square (RMS) fitting] of the ligands. The compounds were fitted to the crystallographic conformation of the template molecule, one of the most active molecules (Fig. 1), and all the aligned molecules of the training set are shown in Fig. 2. Partial atomic charges were calculated using the Del-Re method [17].

Comparative molecular field analysis

The steric and electrostatic CoMFA potential fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å using the Lennard–Jones and the coulomb potentials [18]. The grid box dimensions were determined automatically in such a way that region boundary was extended beyond 4 Å in each direction from coordinates of each molecule. The van der Waals potentials and Coulombic terms, which represent steric and electrostatic fields, respectively, were calculated using tripos force field [19]. An sp^3 -hybridized carbon atom with +1 charge served as probe atom to calculate steric and electrostatic fields. The regression analysis was carried out using the full cross-validated partial least squares (PLS) method [20].

CoMSIA

The CoMSIA [21] descriptors, namely, steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor,

Table 1 Structures and biological activities used in QSAR study

Compound	Set Number	Structure	IC ₅₀ (μM)	pIC ₅₀ (logIC ₅₀)

		Ar		
1	6	4-OCH ₃ -Ph	4.6	5.34
2	4	2,3,4-tri-F-Ph	29	4.54
3	4	2,3-di-F-Ph	39	4.41
4*	4	2,4,5-tri-F-Ph	17	4.77
5	5	3,4-di-F-Ph	8.0	5.10
6*	5	3,5-di-F-Ph	9.4	5.03
7	1	3-CF ₃ -4-CH ₃ -Ph	1.55	5.81
8	2	3,4-di-CH ₃ -Ph	0.35	6.46
9	1	3-CF ₃ -4-Cl-Ph	0.8	6.10
10	4	3-F-4-CF ₃ -Ph	0.077	7.11
11	1	3-CF ₃ -4-Br-Ph	0.45	6.35
12	5	4-F-Ph	19	4.72
13	3	2-naphthyl	0.047	7.33
14	3	2-anthracenyl	0.056	7.25
15	5	3-F-Ph	9.2	5.04
16	1	3-CF ₃ -Ph	14	4.85
17*	1	4-CF ₃ -Ph	0.28	6.55
18	8	3-Cl-Ph	1.4	5.85
19*	8	3-CF ₃ -4-CN-Ph	4.9	5.31
20	5	3-CH ₃ -4-F-Ph	4.6	5.34
21	5	3-F-4-CH ₃ -Ph	0.86	6.07
22*	7	4-C ₆ H ₅ CH ₂ -Ph	2.2	5.66
23	6	4-OCF ₃ -Ph	1.2	5.92
24	6	4-OCHF ₂ -Ph	2.8	5.55
25	8	4-Cl-Ph	1.6	5.80
26	2	3-CH ₃ -Ph	4.9	5.31
27	2	4-CH ₃ -Ph	4.2	5.38
28	7	4-Br-Ph	0.78	6.11

Table 1 (continued)

		R	R1	R2	R3		
29	7	CH3	CH3	H		0.16	6.80
30*	6	CH3	H	H		1.7	5.77
31	7	CH3	H	H		1.2	5.92
32*	3	CH3	H	H		0.33	6.48
33*	2	CH3	H	H		2.0	5.70
34	7	CF3	H	H		0.21	6.70
35	7	C2H5	H	H		0.19	6.72

were generated using a sp³ carbon probe atom with +1.0 charge and a van der Waals radius of 1.4 Å. CoMSIA similarity indices ($A_{F,K}$) between a molecule j and atoms i at a grid point were calculated using Eq. 1 as follows:

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

Fig. 2 Dataset compounds aligned on crystallographic coordinates of compound 14

where q represents the 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 $W_{\text{probe},k}$ is the value of the probe atom.

Five physicochemical properties steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor were evaluated. A Gaussian-type distance dependence was used between the grid point q and each atom i in the molecule. The value of the attenuation factor was set to 0.3. The CoMSIA steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from atomic partial charges, and the hydrophobic fields are derived from atom-based parameters developed by Viswanadhan et al. [22], and hydrogen bond donor and acceptor indices are obtained from a rule based method derived from experimental data.

Partial least square analysis

To quantify the relationship between the structural parameters and the biological activities, the PLS algorithm was used. The CoMFA and CoMSIA descriptors were used as independent variables, and pIC₅₀ values as dependent variables in partial least square regression analysis. PLS was conducted with the standard implementation in the Sybyl 7.1 package. Cross-validation partial least square method of leave-one-out (LOO) was performed to obtain the optimal number of components used in the subsequent analysis. The minimum sigma (column filtering) was set to 1.5 kcal/mol to improve the signal/noise ratio. The optimum number of principle components in the final non-

cross-validated QSAR equations was determined to be that leading to the highest correlation coefficient (r^2) and the lowest standard error in the LOO cross-validated predictions. The non-cross-validation was used in the analysis of CoMFA result and the prediction of the model. Final analysis was performed to calculate conventional r^2 using the optimum number of components obtained from the cross-validation analysis. The result from a cross-validation analysis was expressed as r_{cv}^2 value (Eq. 2):

$$r_{cv}^2 = 1 - \left[\text{PRESS} / \sum (Y - Y_{\text{means}})^2 \right] \quad (2)$$

where PRESS is the sum of the squared deviation between actual (Y) and the predicted activities (Y_{pred}) of training set molecules [$\text{PRESS} = \sum (Y - Y_{\text{pred}})^2$].

To maintain the optimum number of PLS components and minimize the tendency to over fit the data, the number of components corresponding to the lowest PRESS value was used for deriving the final PLS regression models.

The predictive correlation coefficient (r_{pred}^2) based on the test set molecules is computed using formula

$$r_{\text{pred}}^2 = (\text{SD} - \text{PRESS}) / \text{SD} \quad (3)$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activities of training set molecules, and PRESS is the sum of squared deviation between predicted and actual activity for every molecule in test set.

Hierarchical clustering

A 2D distance matrix was calculated on the basis of tanimoto similarity coefficient between every pair of molecules was calculated using open babel [23]. Then, a hierarchical clustering was performed using R statistical package [24] as

shown in Fig. 3. Compounds for training set and test set were selected on the basis of hierarchical clustering.

Training set and test set validation: genetic algorithm based optimization approach

Compounds were classified into eight sets on the basis of hierarchical clustering to ensure the diversity of training and test set.

Thus, each set contains group of molecules having higher tanimoto similarity coefficient with each other. Since compound 33 does not have any sibling, it was assigned to the set containing compound having highest similarity coefficient with compound 33.

Steps followed for the genetic algorithm-based optimization process of training and test set selection are as follows.

Initialization

Initialization generated an initial population of CoMFA models using one randomly selected molecule from each set into test set and rest of the molecules of the same set into training set. The population size was 500.

Repeat

1. Crossover: Roulette wheel selection method was applied to select potentially useful pair of training sets for recombination where probability of being selected for each training set in the population was directly dependent on their q_{LOO}^2 . Single locus point was selected randomly and compounds were swapped between the two test sets corresponding to selected training sets and rearranged the corresponding training set.
2. Mutation: For randomly selected set by a roulette wheel selection method according to the q_{LOO}^2 values, replaced

Fig. 3 Hierarchical clustering tree of dataset compounds

one molecule for test set with randomly selected molecule in the same cluster and rearranged the corresponding training set.

3. Selection: Compared leave-one-out q^2 values of newly created sets with previously generated sets and kept the best models for next generation. After repeatedly performing these steps, the average leave-one-out q^2 values of the individuals in the population increases, as good combination of molecules were discovered and spread through the population.

The step is done until the 200 generations limit is reached.

All these action were performed using sybyl programming language (SPL) scripts.

Molecular dynamics

The protein coordinates contained in the PDB file 3I68, were chosen to start the simulations. All molecular-dynamics simulations were performed using the GROMACS suite of programs (version 4.0.5) [25] using the 43a1 force field. The initial coordinates and topology for HETATOM molecules were constructed with the PRODRG [26] web server. Complexes were placed into cubic box imposing a minimal distance between the solute and the box walls of 10.0 Å and solvated with SPC216 water model. Systems have been neutralized adding the necessary amount of Cl^- ions.

The system was subjected to 500 steps of minimization by steepest descendent method prior to simulations. Following this, 100 ps of position restrained equilibrium run was performed with a force constant of 1,000 KJ/mol Å² on all heavy atoms of the receptor molecule to further equilibrate the medium before starting a full molecular dynamics simulation followed by 2 ns of production run at constant temperature and pressure. Using the leapfrog algorithm in the NPT ensemble, each component, e.g., protein, FMN, H₂O, ORO, inhibitor molecule, and Cl^- , was separately coupled. A cut-off radius of 1.00 nm for short-range repulsive and attractive dispersion interactions, modeled via a Lennard–Jones potential with periodic boundary conditions and the particle mesh Ewald method [27] for long-range electrostatic treatment were used. Constant pressure P and temperature T were maintained by weakly coupling the system to an external bath at 1 bar and 310 K, using the Parrinello–Rahman barostat and Nose–Hoover thermostat, respectively [28]. The system was coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 1 ps, and the isothermal compressibility was $4.5 \times 10^{-5} \text{ bar}^{-1}$. The bond distances and the bond angle of the solvent water were constrained using the SETTLE algorithm [29]. All other bond distances were constrained

using the LINCS algorithm [30], allowing an integration time step of 2 fs.

The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses, gyration radius, and total solvent accessible surface area have been calculated using the GROMACS MD package version 3.1.4 [31] to check the stability and compactness of trajectory. Hydrogen bonds were detected by analyzing the trajectories with the program g_hbond of the GROMACS software.

Results and discussion

The accurate prediction of the bound conformation is a prerequisite for a QSAR model to be reliable. The input ligand conformation was found to have a major impact on the accuracy of the docking results also. Therefore, geometry optimization of all the compounds was performed using quantum mechanics.

Docking

X-ray structures of compounds 13 and 14 reported by Deng and co-workers [32] provide insight into the structural basis of mechanism underlining molecular recognition. To further understand the factors responsible for interaction with the active site residues of PfDHODH and to validate the physical sensibility of docking protocol, active site residues that contribute significantly to the scoring function were extracted. Analyses of docking poses of other structurally diverse compounds in relation to their activities reveal structural requirements for triazolopyrimidine derivatives inhibitory activity consist of a generally planar structure with one or two hydrophobic (aromatic) regions and a polar region (Fig. 4). Hydrogen bonding with residues His185 and Arg265 were found to be crucial for a given set of analogues. In those residues, Gly181, Cys184, His185, Phe188, Leu189, Phe227, Leu531, and Val 532 were found to be the most important residues in the active site for VDW interactions. Phe188 also functions by forming π – π interaction with ligand, while the other residues define the shape and size of the hydrophobic cavity. The ring of Phe227 is almost perpendicular to the phenyl ring of the ligand and is involved in the formation of a blocking wall to prevent the ligand ring from moving away from the position where it forms a π – π interaction with the Phe188 ring. Residues Asp169, Glu182, His185, Arg265, and Leu531 are supposed to be important in providing electrostatic interactions in the active site. Gly181, Cys184, His185, Phe188, Leu531, and Val532 are probably helpful in enhancing the activity of ligands with polar groups oriented in the cavity area. Analysis of docking results also indicate the existence of repulsive interactions due to the ortho-fluoro phenyl

Fig. 4 Binding mode conformation of docked compound 14 (*magenta color*) relative to its cocrystalised conformation (*green color*) in the active site of pfDHODH. The hydrogen bonds formed between docked conformation and active site residues (with His185 and Arg265) are shown in *red*

substituent pointing into an electron-rich environment made up of Phe188 and Phe227. Moreover p-substituted halogens and aryl substituents lie in pocket composed of Phe 227, Leu531, Phe 188, and Ile 237 possess both hydrophobic and fluorophilic characteristics.

Training and test set selection

To exclude homogeneous data from training and test set, clustering was performed. It provides an assurance that all the chemical classes are represented in the training set. Otherwise, there may be an apparent risk that small clusters with few members will not be represented in the final training set. This also leads to a test series of compounds in which all major structural and chemical properties are symmetrically varied at the same time.

Effective descriptor or variable selection is an important step in the QSAR modeling process. To achieve this goal, selection of training and test sets was manipulated based on genetic algorithm to maximize the predictive capability of the model being published. The process is based on the assumption that training set has covered all the available structure space and a molecule that is structurally very

similar to the training set molecules will be predicted well because the model has captured features that are common to the training set molecules and is able to find them in the new molecule.

To evaluate the performance of the GA analysis, a total of 200 GA runs were performed. The best CoMFA-based model in each GA run was constructed for comparison. Out of the 200 models, the top 20 models were selected for further analysis. Overall, these results indicate that models are statistically comparable.

The predictive power of the 3D QSAR models was evaluated by predicting the activities of the eight compounds belonging to the test set. The predictive ability of the models is expressed by the predictive r^2 value (r^2_{pred}). All 3D-QSAR statistical results are summarized in Table 2. At eight number of components, CoMFA model has cross-validated coefficient (leave-one-out) q^2_{LOO} of 0.841, q^2 (cross-validated) at tenfold of 0.818 and non-cross-validated r^2 of 0.99 with standard error of estimate (SEE) of 0.033.

The two models were further used for test set which gives r^2_{pred} of 0.88 for CoMFA. The CoMSIA models were developed for the top 20 set of models obtained from GA

Table 2 Summary of the CoMFA and CoMSIA statistical results for the training set molecules

	CoMFA	CoMSIA
Q_2 (Leave-one-out)	0.841	0.757
q^2 (cross-validated)	0.818	0.653
r^2	0.99	0.943
SEE ^a	0.033	0.212
N_c^b	8	4
Field contribution (%)		
Steric	0.785	0.540
Electrostatic	0.215	0.202
HB acceptor		0.257

^a Standard error of estimate^b Optimum number of components

optimization process. Because the five different descriptor fields are not totally independent of each other and such dependencies may reduced the statistical significance and predictivity of models, possible combinations of different fields with positive value of r^2_{pred} for test set were analyzed further (Fig. 5). The combination of steric (S), electrostatic (E), and H-bond acceptor (A) was considered for further analysis as it provides optimal values of statistical parameters, q^2_{LOO} and r^2_{pred} . The CoMSIA model was reported with a q^2_{LOO} of 0.757, r^2_{nv} of 0.943, q^2 (cross-validated) at tenfold of 0.653, and r^2_{pred} of 0.466.

Statistically, steric, electrostatic, and H-bond effects account for 54.0%, 20.2%, and 25.7%, respectively. According to the fact that q^2 and r^2_{pred} is usually used as a measure of 3D QSAR quality, therefore, taking all statistical results into account, the CoMFA model in terms of higher q^2 and

Fig. 5 Results of the possible CoMSIA field combinations (*S* steric, *E* electrostatic, *H* hydrophobic, *D* H-bond donor, *A* H-bond acceptor) with their respective q^2 values (LOO cross-validation using the PLS method) and r^2_{pred} obtained for test set

r^2_{pred} values is more explanatory than CoMSIA model for the chosen set of training set compounds. The test set points are placed above and below the correlation line of CoMFA and CoMSIA models (Fig. 6), indicating that the prediction ability of CoMFA model is correct (Table 3).

Contour map analysis

The contour maps derived from the CoMFA and CoMSIA PLS model have permitted an understanding of the steric and electrostatic requirements for ligand binding. The results obtained from CoMFA and CoMSIA PLS models were graphically interpreted through the stdev*coefficient color-coded contour maps (Fig. 7a and b) obtained after contour analysis for deriving relationship between molecular field differences of a set of 35 triazolopyrimidine derivative molecules and differences in their biological activities. In case of CoMFA contour model, the electrostatic map is represented by red and blue contours, where red contour indicates enhanced biological activity with increased negative charge, and the blue contour indicates enhanced biological activity with increased positive charge. Similarly, the steric contour is represented by green and yellow colors, where green contours indicate higher activity with sterically bulky group, while the yellow contours indicate decrease in

Fig. 6 Graphs of experimental value vs. predicted values for training and test set compounds. **a** CoMFA, **b** CoMSIA (square training set; triangle test set)

Table 3 Actual and CoMFA and CoMSIA-based predicted activities of triazolopyrimidine analogues

Compound	Actual pIC50	CoMFA		CoMSIA	
		Predicted pIC50	residual	Predicted pIC50	residual
1	5.34	5.357	−0.02	5.256	0.08
2	4.54	4.467	0.07	4.614	−0.08
3	4.41	4.444	−0.03	4.567	−0.16
4 ^a	4.77	4.486	0.28	4.664	0.11
5	5.1	5.147	−0.05	5.144	−0.05
6 ^a	5.03	4.552	0.48	4.55	0.48
7	5.81	5.816	−0.01	5.758	0.05
8	6.46	6.441	0.02	6.344	0.11
9	6.1	6.143	−0.05	5.732	0.37
10	7.11	7.113	0	7.397	−0.28
11	6.35	6.346	0	6.377	−0.03
12	4.72	4.708	0.01	4.818	−0.1
13	7.33	7.314	0.01	7.249	0.08
14	7.25	7.25	0	7.286	−0.03
15	5.04	5.034	0	4.764	0.27
16	4.85	4.829	0.02	5.144	−0.29
17 ^a	6.55	6.639	−0.09	6.243	0.31
18	5.85	5.815	0.04	5.882	−0.03
19 ^a	5.31	5.661	−0.35	5.563	−0.25
20	5.34	5.349	−0.01	5.496	−0.16
21	6.07	6.076	−0.01	5.798	0.27
22 ^a	5.66	5.875	−0.22	6.816	−1.16
23	5.92	5.914	0.01	6.086	−0.16
24	5.55	5.578	−0.02	5.516	0.04
25	5.8	5.804	−0.01	5.919	−0.12
26	5.31	5.339	−0.03	5.55	−0.24
27	5.38	5.348	0.03	5.501	−0.12
28	6.11	6.112	0	5.576	0.53
29	6.8	6.774	0.02	6.686	0.11
30 ^a	5.77	5.667	0.1	5.914	−0.14
31	5.92	5.913	0.01	6.079	−0.16
32 ^a	6.48	6.223	0.26	6.066	0.42
33 ^a	5.7	5.962	−0.26	5.89	−0.19
34	6.7	6.736	−0.04	6.651	0.05
35	6.72	6.686	0.04	6.663	0.06

^aTest set compound

activity with increase in bulk. The total field contribution provided by electrostatic field is 21.5%, and steric field is 78.5% for CoMFA. Highly active compound was embedded in the CoMFA and CoMSIA contour maps to demonstrate its affinity for the steric and electrostatic regions of inhibitors.

CoMSIA, a distance dependent Gaussian-type functional form, takes hydrophobic, hydrogen bond donor, and acceptor components also into consideration with steric and electrostatic fields for building models. In CoMSIA methods, the steric fields are represented by green- and yellow-colored contours (green, bulky substitution favored; yellow,

bulky substitution disfavored); the electrostatic fields are indicated by red- and blue colored contours (blue, electro-positive group favored; red, electronegative group favored); In case of CoMSIA, in addition to steric and electrostatic fields, the hydrogen bond acceptor fields are denoted by magenta and cyan contours (magenta, favored; cyan, disfavored).

From the CoMFA and CoMSIA contour map analysis of a given training set, it is clear that variation around phenyl ring is more desirable. CoMFA steric and electrostatic field contours are shown in Fig. 7a. Single prominent green contour present in the vicinity of the seven and eight

Fig. 7 **a** CoMFA steric and electrostatic contours displayed with most potent compound in the active site. **b** CoMSIA steric and electrostatic and hydrogen bond acceptor contours displayed with most potent compound in the active site

positions of the naphthyl ring indicate that generally steric bulks are favored at these sites. The good inhibitory potency of compounds 13 and 14 is due to orientation of the benzene ring toward the sterically favored regions. The electrostatic contours of CoMFA show prominent red regions surrounding the naphthyl ring, indicating that incorporation of electron-rich substituents would enhance the activity. Red contours in the vicinity of both of the side chains substituted at C-3 and C-6 of the ring have been observed in compounds 9, 10, 11, and 17 with remarkable activity while in the case of compounds 2, 3, and 12 orientation of electronegative group towards the blue contours makes these compounds poor inhibitors. This highlights the requirement of electronegative substituents at proper place with proper orientation as also indicated by Ojha et al. [6] that fluoro-substituents at ortho-position show lower range of activity while hydrophobic substituents at m- and p-positions show better potency for this class of compounds.

Information obtained from CoMSIA contour maps is almost similar to that obtained from CoMFA contour maps with respect to steric and electrostatic effects except larger size of green sterically favorable contour in case of CoMSIA, which mislead to high predictive activity of compound 6 by CoMSIA model. In addition, hydrogen bond acceptor

contour shown in Fig. 7b indicate that the hydrogen bond acceptor favorable magenta region is also found.

Molecular dynamics

In recent years, the role of halogens especially fluorine in medicinal chemistry and drug design has been studied extensively [33–36], as fluorine show quite distinct qualities than other halogens due to its high electronegativity and low polarisability. According to the SAR fundamental theory, similar structures should have similar activities, but in the present dataset, compounds 10 and 14, two highly potent inhibitors, are somewhat dissimilar as also shown by their presence in two distinct clusters (Fig. 3).

To further rationalize their high activities despite low tanimoto similarity coefficient between them, molecular dynamic studies of these two compounds were performed in order to study the stability of molecular interactions of ligands in solution over time with the active site residues observed in molecular docking studies as many penalty terms (e.g., steric and electrostatic clash, internal ligand strain) are not easy to correctly parameterize in docking studies. In particular, entropy and desolvation are difficult to treat accurately even within a rigorous molecular

Fig. 8 RMSD of backbone (a). Of protein complex with compound 10 (gray line) (b). Of protein complex with compound 14 (black color) over trajectory

mechanics formalism [37]. On the other hand, molecular simulation constitutes a useful tool to elucidate the conformation of the ligand in protein.

Two nanosecond molecular dynamics calculations were performed on the PfDHODH complexed with compounds 10 and 14 separately. The main chain RMSDs were calculated, for the trajectories of the two protein complexes, from the starting structures as a function of time to evaluate the conformational flexibility of the system (Fig. 8). Although the RMSDs of both the systems reached conformational equilibrium within the first 500 ps and showed a plateau for the rest of the simulation, which confirm the protein stability over the entire trajectory chosen for the analysis, all the analyses were carried out after discarding the first 700 ps.

Smaller RMSF values of ligands atoms showed tight interaction between active site residues of receptor and inhibitor molecules (Fig. 9). Hydrophobic residues like Leu197, Ile237, Leu240, Leu531, and Met536 were found to show greater fluctuation in case of receptor molecule complexed with compound 10 as compared to that of compound 14 over time, which highlights the lack of hydrophobic interaction between residues and ligand atom due to presence of smaller CF₃ group. In our study, phenyl ring of compound 10 attained almost orthogonal conformation from initial docking conformation after MD run, while no such deviation was seen for co-crystallized conformation of compound 14, indicating the limitation of docking programs for insightful study of molecular recognition processes. Moreover, MD simulation indicated the possible presence of orthogonal multi-polar non bonding interactions between m-fluoro substituent and fluophilic C=O group of Leu531 support the observations of Ojha et al. [6] for favorability of fluoro group at meta position but not favorable at ortho position of phenyl ring whereas trifluoromethyl group are

supposed to attribute to the larger hydrophobic surface area upon binding.

The dynamics of the hydrogen bonds of compounds 10 and 14 were quite different. The hydrogen bond between the atom ND1 of residue His185 and the atom N1 of compound 14 broke and reformed frequently several times as compared to hydrogen bond between the atom ND1 of residue His185 and atom N1 of compound 10 during the trajectory, which is found to be more consistent. Similar trend was found in the two complexes with the NH1 of Arg265 and N5 of the ligands. It can be concluded, therefore, the occurrence of hydrogen bonds over time with His185 and Arg265 of compound 10 is found to be stronger as compared to that of with compound 14. This indicates that the presence of electron withdrawing group (–CF₃) increases the polarity of compounds and so the electrostatic interaction with nearby active site residues. Whereas between His185 and Arg265, hydrogen bonding with Arg265 was found to be stronger as compared to that of with His265 with higher hydrogen-bond lifetime and number of times of occurrence of hydrogen bond over the trajectory. It may indicate that the gain in the binding affinity upon the replacement by halogen groups may not arise from halogen/fluorine binding only, and the properties of fluorine could be effectively exploited to selectively enhance the ligand affinity in structure-based design.

Conclusions

In the present paper, we have used a novel selection method of training set for CoMFA modeling. We applied this approach to the data set of the PfDHODH inhibitor molecules. Our selection method gave simpler and significantly improved 3D QSAR model equations in lesser time compared

Fig. 9 Average root mean square fluctuation of a compound 10(gray color) and compound 14 (black color)

with those from the conventional CoMFA. The structural requirements for the PfDHODH inhibitor molecules could be easily estimated from the simplified 3D coefficient contour maps of the final CoMFA model. These analyses guarantee that both training and test sets represent the structural diversity and cover the whole data set potency and selectivity space, rendering the data set appropriate for the purpose of QSAR model development. It is important to note that the same training and test sets were employed for all 3D-QSAR analyses.

The results obtained from molecular dynamics simulations indicate that both the protein complexes display a stable structure that is fully maintained over the entire simulation time. Consensus pattern in the two simulation results explain the validity of the docking parameters. Docking and MD simulation results agree well with QSAR contours; hence, they successfully complement each other. Information obtained from this study would be helpful for understanding PfDHODH–ligand relationship and therefore designing of more potent inhibitors targeting this enzyme.

Acknowledgments This work was supported by the grants from Council of Scientific and Industrial Research (CSIR-India) funded network project NWP0034 (Validation of identified screening models and development of new alternative models for evaluation of new drug entities). PS thanks CSIR for a fellowship. This manuscript is a CDRI communication no. 8189.

References

- Nicolotti O, Miscioscia TF, Carotti A, Leonetti F, Carotti A (2008) J Chem Inf Model 48:1211–1226
- Doweyko AM (2004) J Comput Aided Mol Des 18:587–596
- Guo J, Hurley MM, Wright JB, Lushington GH (2004) J Med Chem 47:5492–5500
- Patel V, Booker M, Kramer M, Ross L, Celatka CA, Kennedy LM, Dvorin JD, Duraisingh MT, Sliz P, Wirth DF, Clardy J (2008) J Biol Chem 283:35078–35085
- Heikkilä T, Thirumalaiah S, Davies M, Parsons MR, McConkey AG, Fishwick CW, Johnson AP (2006) Bioorg Med Chem Lett 16:88–92
- Ojha PK, Roy K (2010) Eur J Med Chem 45(10):4645–4656
- Leonard JT, Kunal R (2006) QSAR Comb Sci 25:235–251
- Yuan H, Petukhov PA (2006) Bioorg Med Chem Lett 16:6267–6272
- Gujjar R, Marwaha A, El Mazouni F, White J, White KL, Creason S, Shackleford DM, Baldwin J, Charman WN, Buckner FS, Charman S, Rathod PK, Phillips MA (2009) J Med Chem 52:1864–1872
- Phillips MA, Gujjar R, Malmquist NA, White J, El Mazouni F, Baldwin J, Rathod PK (2008) J Med Chem 51:3649–3653
- SYBYL Molecular Modeling System Version 7.1 (2005) Tripos Inc., St. Louis, MO, USA
- Michael WS, Kim KB, Jerry AB, Steven TE, Mark SG, Jan HJ, Shiro K, Nikita M, Kiet AN, Shujun S, Theresa LW, Michel D Jr, John AM (1993) J Comput Chem 14:1347–1363
- Berendsen HJC, van der Spoel D, van Drunen R (1995) Comput Phys Commun 91:43–56
- van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. VdF: Hochschulverlag AG an der ETH Zürich and BIOMOS b.v, Zürich
- Rarey M, Kramer B, Lengauer T, Klebe G (1996) J Mol Biol 261:470–489
- Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) J Comput Aided Mol Des 11:425–445
- Delre G, Pullman B, Yonezawa T (1963) Biochim Biophys Acta 75:153–182
- Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
- Matthew C, Richard DC III, Van Nicole O (1989) J Comput Chem 10:982–1012
- Bush BL Jr, Nachbar RB (1993) J Comput Aided Mol Des 7:587–619
- Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130–4146
- Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) J Chem Inf Comput Sci 29:163–172
- OpenBabel v.2.2.0. <http://openbabel.org>
- The R Project for Statistical Computing. <http://www.r-project.org/>
- Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) J Chem Theory Comput 4:435–447
- Schüttelkopf AW, van Aalten DM (2004) Acta Crystallogr D: Biol Crystallogr 60:1355–1363
- Darden T, York D, Pedersen L (1993) J Chem Phys 98:10089–10092
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684
- Shuichi M, Peter AK (1992) J Comput Chem 13:952–962
- Berk H, Henk B, Herman JCB, Johannes GEMF (1997) J Comput Chem 18:1463–1472
- Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341
- Deng X, Gujjar R, El Mazouni F, Kaminsky W, Malmquist NA, Goldsmith EJ, Rathod PK, Phillips MA (2009) Structural plasticity of malaria dihydroorotate dehydrogenase allows selective binding of diverse chemical scaffolds. J Biol Chem 284(39): 26999–27009
- Hagmann WK (2008) The many roles for fluorine in medicinal chemistry. J Med Chem 51(15):4359–4369
- Müller K, Faeh C, Fo D (2007) Fluorine in pharmaceuticals: looking beyond intuition. Science 317(5846):1881–1886
- Voth AR, Khuu P, Oishi K, Ho PS (2009) Halogen bonds as orthogonal molecular interactions to hydrogen bonds. Nat Chem 1(1):74–79
- Lu Y, Wang Y, Zhu W (2010) Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design. Phys Chem Chem Phys 12(18):4543–4551
- Waszkowycz B, Clark DE, Gancia E (2011) Outstanding challenges in protein–ligand docking and structure-based virtual screening. Wiley Interdiscip Rev Comput Mol Sci 1(2):229–259