ORIGINAL PAPER

New molecular scaffolds for the design of *Mycobacterium tuberculosis* type II dehydroquinase inhibitors identified using ligand and receptor based virtual screening

Ashutosh Kumar · Mohammad Imran Siddiqi · Stanislav Miertus

Abstract Using ligand and receptor based virtual screening approaches we have identified potential virtual screening hits targeting type II dehydroquinase from *Mycobacterium tuberculosis*, an effective and validated *anti*-mycobacterial target. Initially, we applied a virtual screening workflow based on a combination of 2D structural fingerprints, 3D pharmacophore and molecular docking to identify compounds that rigidly match specific aspects of ligand bioactive conformation. Subsequently, the resulting compounds were ranked and prioritized using receptor interaction fingerprint based scoring and quantitative structure activity relationship model developed using already known actives. The virtual screening hits prioritized belong to several classes of molecular scaffolds with several available substitution positions that could allow chemical modification to enhance binding affinity. Finally, identified hits may be useful to a medicinal chemist or combinatorial chemist to pick up the new molecular starting points for medicinal chemistry optimization for the design of novel type II dehydroquinase inhibitors.

Keywords Molecular docking · Molecular fingerprints · 3D pharmacophore · QSAR · Type II dehydroquinase

A. Kumar · M. Imran Siddiqi
Molecular and Structural Biology Division,
Central Drug Research Institute,
Lucknow 226 001, India

M. Imran Siddiqi (✉) · S. Miertus (✉)
International Centre for Science and High Technology, UNIDO,
AREA Science Park,
Trieste 34012, Italy
e-mail: imsiddiqi@yahoo.com
e-mail:stanislav.miertus@ics.trieste.it

Introduction

The shikimate pathway is the biosynthetic route to the aromatic amino acids and other important aromatic metabolites in plants, bacteria, fungi, and apicomplexan parasites [1–3]. The pathway is absent in mammals, making the corresponding enzymes attractive targets for the development of new herbicides and antimicrobial agents [1]. Dehydroquinase (3-dehydroquinate dehydratase, EC 4.2.1.10), the third enzyme of the shikimate pathway, catalyses the conversion of 3-dehydroquinate to 3-dehydroshikimate. There are two forms of dehydroquinase, type I and type II, which appear to have arisen by convergent evolution. These enzymes are structurally distinct and catalyze the same overall transformation by very different mechanisms [4]. The type I dehydroquinases (for example, from *Escherichia coli*) are dimeric proteins with a 26–28 kDa subunit catalyze the *syn* dehydration of 3-dehydroquinate through the initial formation of a Schiff base with a conserved lysine residue [5, 6]. In contrast, type II enzymes (for example, from *Streptomyces coelicolor* and *Mycobacterium tuberculosis*) are dodecamers of 12–18 kDa subunits and catalyze the *anti* elimination of water *via* an enolate intermediate [7–9]. Since only type II dehydroquinases are present in *M. tuberculosis* [10] therefore specific inhibitors of this enzyme have therapeutic potential.

The elimination mechanism of the type II dehydroquinases proceeds *via* an enol intermediate characterized by ring-flattening due to the π-system formed between C2 and C3 and by the increased H-bonding of the enol oxygen formed, so the compounds that mimicked the enol intermediate were likely to show inhibitory properties. This feature has been exploited for the design of the first generation of inhibitors and was first studied by Abell et al., who reported a number of analogues of 3-dehydroquinic acid as compet-

itive reversible inhibitors of *M. tuberculosis* type II dehydroquinase [11–14]. 2, 3-Dehydroquinic acid was initially designed to mimic the flattening of the ring in the enol intermediate [11] and subsequently, in the search for more accurate analogues, proton H3 of 2, 3-Dehydroquinic acid was replaced by a fluorine atom in order to mimic the high electron density generated at C3 during the course of the reaction [12–14].

Co-crystallization of anhydroquinate analogue with *S. coelicolor* type II dehydroquinase revealed a second binding pocket adjacent to the active site which was adventitiously occupied by a glycerol molecule from the enzyme storage buffer [15]. This binding pocket is located underneath a flexible loop, containing the catalytically important tyrosine and arginine residues. The identification of this second binding pocket inspired the design of bifunctional compounds with a quinate core and side chains from C-3 to reach into the second binding site [16–18]. Prazeres et al. showed that incorporation of aryl groups bearing electron withdrawing substitutions in C-3 positions of quinate core markedly increased the inhibitory potency of this enol mimic against *M. tuberculosis* type II dehydroquinase [19]. In fact these studies led to the discovery of the most potent known inhibitor against any type II dehydroquinase, the 3-nitrophenyl derivative with Ki value of 54nM against *M. tuberculosis* type II dehydroquinase [20]. Despite these and other known inhibitors, more structurally diverse inhibitors of *M. tuberculosis* type II dehydroquinase need to be discovered for improving the understanding of the biological function of *M. tuberculosis* type II dehydroquinase and to discover its potential therapeutic indications.

The identification of new lead candidates is a crucial task in the early phase of drug discovery. The general goal is to select a small number of compounds with desired properties (*e.g.*, bioactivity against a drug target) from hypothetically available screening compounds [21]. Chemical space is vast −the number of synthetically accessible organic molecules has been estimated to be in the range of 10^{60}−10^{100} [22–24]. It is evident that exhaustive screening of such a large number of substances is by no means possible. Advances in high-throughput screening (HTS) and parallel synthesis since the early 1990s have provided a valuable tool for standard pharmaceutical research, [25] and large compound libraries can be synthesized in a combinatorial fashion and screened with help of robotics [26]. HTS campaigns demand a considerable financial effort and do not always yield many validated hits [27–29]. Alternatively, computational approaches like similarity searches [30], pharmacophore searching, [31] molecular docking, [32] QSAR methods, [33] and de novo design [34] concentrate on "cherry-picking" of selected compounds−often only tens to hundreds−with predicted desired activity. A complementary

approach for computational lead identification is targeted library design [35–37]. In this context, we have decided that the sequential combination of ligand similarity, pharmacophoric features, molecular docking strategies with interaction fingerprints and QSAR in a single workflow could be useful for this purpose. The fact that these methods, when used individually, focus only on one part of the structural information available has recently prompted the development of hybrid ligand and receptor based virtual screening methods. These hybrid methods aim at fully exploiting all the structural information present in ligand-bound protein structures, both from the protein and ligand perspective. 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. 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. Such a combined approach uses ligand-based methods to identify compounds with features important for the target property. Structure-based techniques are used to ensure that the shape, size and energetic interaction potential of the putative ligands complement that of the target protein. In principle, a strategy combining 2D and 3D ligand similarity and structural information of protein should provide more accurate prediction about ligand binding modes. 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 [36]. Therefore, here in this study iterative search strategies are employed to perform "smart" sampling of screening compounds that incorporate as much available knowledge about the target of interest as possible, in an effort to make lead discovery more productive, efficient, and cost-effective. The result of targeted library design is a small molecular library enriched with desired compounds that are focused toward a specific biological target and may be useful to a medicinal chemist or combinatorial chemist to pick up the new molecular scaffolds for medicinal chemistry optimization for the design of novel inhibitors.

The aim of our study was to screen a possibly diverse set of small molecular structures, in order to obtain scaffolds capable of fitting into *M. tuberculosis* type II dehydroquinase active site. The resulting candidates for novel compounds should hopefully benefit from specificity provided by the quinate derivatives as well as affinity resulting from reproducing the key interactions. We performed sequential virtual screening by means of a 2D similarity search, 3D

pharmacophore search and virtual screening hit prioritization using molecular docking, and quantitative structure activity relationship. Such a method allows combining receptor-based and ligand-based approaches, thus utilizing most of the currently available structural data. The novelty of this approach lies in the sequential combination of ligand (2D similarity and 3D pharmacophore) and structure based (molecular docking and interaction fingerprints) virtual screening facilitated through the prediction of hit compounds activities using quantitative structure activity relationship for screening against the type II dehydroquinase from *M. tuberculosis*. Another important feature of this computational approach is the enhancement of its predictive power by incorporating an exhaustive consensus scoring process using a rank-by-rank strategy in order to provide a list of suitable compounds to be considered for bioscreening. Our results show that ligand-based and structure-based approaches can be used to discriminate actives from inactives in a retrospective virtual analysis. We believe that these approaches can be integrated in prospective virtual analysis to select candidates for the rational design of *M. tuberculosis* type II dehydroquinase inhibitors.

Materials and methods

Dataset

A total of 45 *M. tuberculosis* type II dehydroquinase inhibitors and their affinities were collected from the literature [19, 20, 38–40], given as K_i values (Table 1). To evaluate the performance of ligand and receptor based virtual screening; presumably inactive compounds or decoys (500 in number) were prepared from Maybridge small molecule database using the method described by Huang et al. [41]. We have used similarity and physical property analysis to ensure that decoys have similar physical properties (*e.g.*, molecular weight, number of hydrogen bond donors/acceptors, number of rotatable bonds and LogP) to the actives. The ratio of active to inactive compounds was 1:10 which is in the range of ratios chosen in similar studies, *e.g.*, a ratio of 1:17 was chosen by Venhorst et al. [42] and a ratio of 1:19 was chosen by Evers et al. [43]. As such, this data is expected to allow a bias-free assessment of virtual screening methods. Noteworthy, these molecules were only assumed to be inactive with no experimental evidence that there is no cross reactivity with *M. tuberculosis* type II dehydroquinase.

2D fingerprint based similarity search

The 2D similarity searches were performed with the software MOE [44] using two-point (typed graph distances

(TGD)) and three-point (typed graph triangles (TGT) pharmacophore-based fingerprints, all calculated from a 2D molecular graph. Each atom was given a type among donor, acceptor, polar, anion, cation, or hydrophobe for the calculation of TGD and TGT. Subsequently, pairs (two-point fingerprint) or triplets (three-point fingerprints) of types were formed by graph distances and coded as sparse features in a fingerprint. The Tanimoto coefficient was used as the similarity metric [45]. The Tanimoto coefficient between two molecules described by a 2D fingerprint is calculated using the following expression:

$$T_{ab} = \frac{c}{a + b - c} \tag{1}$$

where "*c*" is the number of bits common to the two molecules, and "*a*" and "*b*" denote the number of bits set in each of the two fingerprints.

3D Pharmacophore search

A 3D pharmacophore query was defined on the basis of the structural features of the potent *M. tuberculosis* Type II dehydroquinase inhibitor 3-nitrophenyl derivative of quinic acid (Compound **28** in Table 1), by using the putative bound conformation of the ligand derived from the docking. The 3D pharmacophore search was performed by using the Unity flexible search protocol as implemented in Sybyl 7.1, with all options set as default [46]. In the Unity search, the conformations of the screening database were generated on the fly by means of the Directed Tweak method [47].
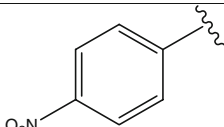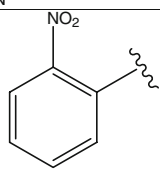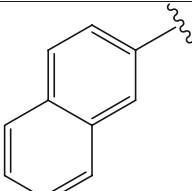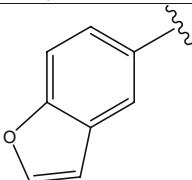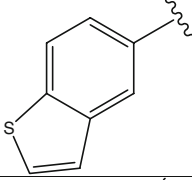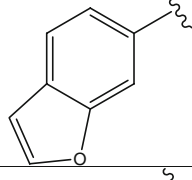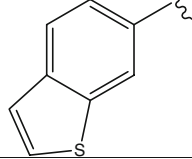
Molecular docking and scoring

The hits obtained from 3D pharmacophore search were docked into Type II dehydroquinase binding pocket in complex with inhibitor (PDB entry code 1H0R) using the FlexX program interfaced with Sybyl7.1 [46]. Standard parameters of the FlexX program as implemented in SYBYL7.1 were used during docking. To further evaluate the docking results, the G_Score [48], PMF_Score [49], D_Score [50] and ChemScore [51] values were estimated using the CScore module of Sybyl7.1 [46]. As CScore is a consensus scoring function, the different scoring functions in it provide multiple approaches to evaluate ligand–receptor interactions and such different scores are expected to better aid in prioritization.

Receptor interaction fingerprints

Receptor interaction fingerprints were generated using the method developed by Rognan et al. [52] from docked poses of the virtual screening hits and active site coordinates taking 3-nitrophenyl derivative of quinic acid (Compound
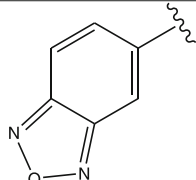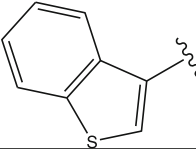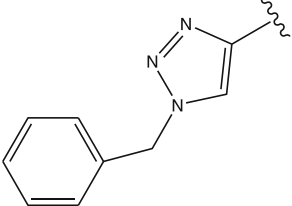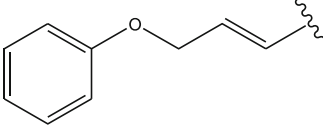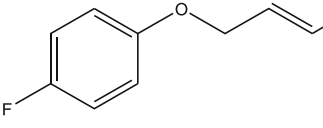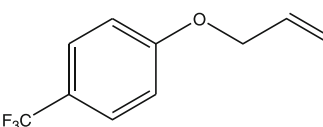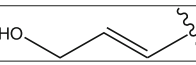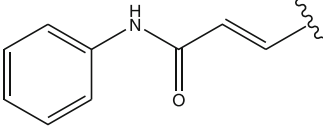
Table 1 Structure, biological activities and predicted activities of *M. tuberculosis* Type II dehydroquinase inhibitors

Compound. No.	R	Ki(µM)	-logKi	pKi	Compound Ref. No.
1	H	200	3.699	3.541	19
2	F	10	5.00	5.080	19
3		6.5	5.187	5.730	19
4		109	3.963	3.968	19
5		1.2	5.921	5.827	19
6*		2.0	5.699	5.675	19
7*		2.85	5.545	5.327	19
8		0.97	6.013	6.207	19
9		0.85	6.071	5.633	19

28) as reference ligand using a C++ library and executables available from http://bioinfo-pharma.u-strasbg.fr. In the present work, only the first seven bits (Hydrophobic interactions, aromatic face to face, aromatic face to edge, hydrogen bond acceptor, hydrogen bond donor, positively charged and negatively charged) which correspond to the most frequent protein-ligand interactions are calculated. The distance between two interaction fingerprints was

Table 1 (continued)

10*		1.95	5.710	5.753	19
11		13.2	4.879	5.144	19
12		3.75	5.426	5.498	19
13		0.14	6.854	6.848	36
14*		0.12	6.921	7.132	36
15		4.3	5.367	5.120	36
16*		17	4.770	4.562	36
17		2.3	5.638	5.391	36

calculated using a Tanimoto similarity coefficient (Tc) as follows:
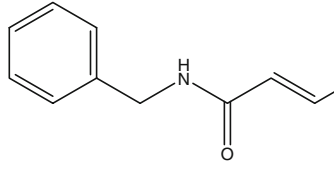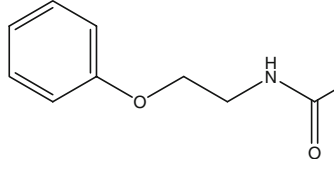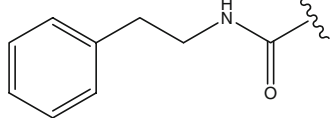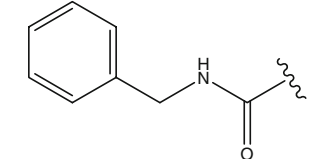
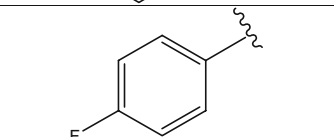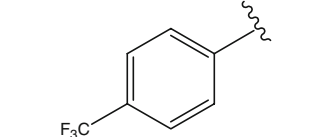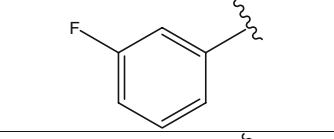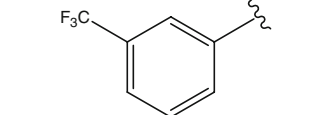$$T_c = \frac{|A \cap B|}{|A \cup B|} \qquad (2)$$
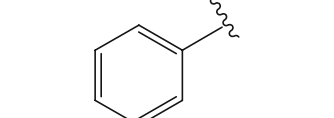
where $A \cap B$ is the number of switched-on bits common to fingerprints A and B and $A \cup B$ is the sum of switched-on bits in fingerprints A and B.

Quantitative structure activity relationships

A total of 45 *M. tuberculosis* Type II dehydroquinase inhibitors and their affinities reported by different groups but evaluated under the similar conditions were collected from the literature [19, 20, 38–40] and used in the present study to derive and validate the GA-MLR QSAR models using MOE Version 2007 [44]. The K_i values were converted to the corresponding pK_i (-log K_i) in which higher values indicate exponentially greater potency. The pKi values were then used as dependent variables in QSAR analysis. A rationally designed training set of 35 molecules was used for the derivation of the models and the test set of 10 molecules indicated by asterisk was used for their validation (Table 1). The rational selection of 35 molecules as the training set and 10 molecules as test set was accomplished using the diverse subset tool of MOE using
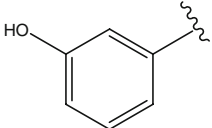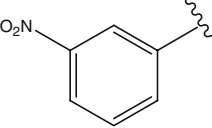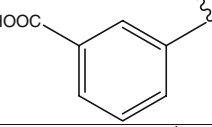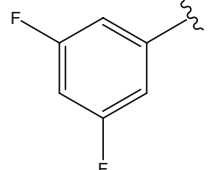
Table 1 (continued)
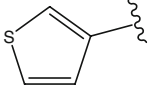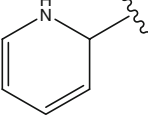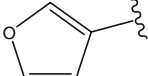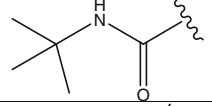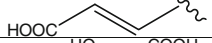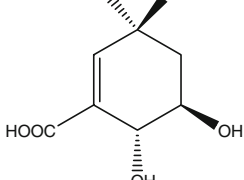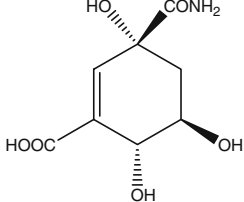
18		20	4.699	4.725	36
19		49	4.310	4.372	36
20*		8.7	5.060	5.434	36
21		0.91	6.041	5.775	36
22		1.5	5.824	5.730	20
23		1.8	5.745	6.028	20
24		17	4.770	4.801	20
25		1.5	5.824	5.647	20
26		2.125	5.673	5.887	20

MACCS keys and Tanimoto similarity coefficient. Both 2D and 3D molecular descriptors were calculated using MOE Version 2007 [44]. 2D descriptors include physical properties, subdivided surface areas, atom counts and bond counts, Kier and Hall connectivity [53–55] and kappa shape indices; [56, 57] adjacency and distance matrix descriptors, [58–62] pharmacophore feature descriptors, and PEOE partial charge descriptors [63]. 3D molecular descriptors include potential energy descriptors, surface area, volume and shape descriptors, and conformation-dependent charge descriptors [64]. Other descriptors include Lipinskii's descriptors, Oprea descriptors, reactive functional groups, Xu and Stevenson's drug like index descriptors, MACCS descriptors and E-state descriptors. All of these descriptors were calculated using automated MOE-SVL scripts.

Thereafter, genetic algorithm (GA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated

Table 1 (continued)

No.	Structure				
27		95	4.022	4.208	20
28		0.054	7.268	6.754	20
29		150	3.824	3.828	20
30		2.44	5.613	5.587	20
31*		0.59	6.229	6.009	20
32		45	4.347	4.288	20
33		0.83	6.081	6.154	20
34		27	4.569	4.257	37
35		105	3.979	4.011	37
36*		0.94	6.027	5.893	37
37		16	4.796	4.716	37

descriptors, *i.e.*, multiple linear regression modeling (MLR). The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF) [65]. The MLR method is a simple and classical regression method, which can provide explicit equations. In the current work, the models were built using the simple MLR method with the selected variables from GA, called GA-MLR. In the present studies the genetic algorithm has been implemented through the QuaSAR-Evolution algorithm provided in MOE. [The QuaSAR-evolution script was downloaded from the SVL exchange, http://svl.chemcomp.com.] This algorithm supports multiple fitness evaluation scores: the LOF, cross-

Table 1 (continued)

38		13	4.886	4.704	38
39		8	5.097	5.081	38
40		15	4.824	4.839	38
41		1200	2.921	3.005	38
42*		700	3.155	3.122	38
43		20	4.699	4.812	20
44		20	4.699	4.741	38
45*		700	3.155	3.127	38

* Test Set Compounds

validated correlation coefficient (q^2), correlation coefficient r^2 and the adjusted r^2. Nevertheless, in view of the superiority of the LOF fitness function in comparison to the above listed functions, it was alone used for the implementation of the GA.

Results and discussion

The general strategy for the ligand and structure based virtual screening pursued in the present study is presented in Fig. 1. In summary, using 2D structural fingerprints we initially analyzed ZINC database [66] for similarity to a highly active Type II dehydroquinase inhibitors (Compound 13, 28 and 31 in Table 1). Subsequently, compounds selected in this similarity search were subjected to 3D pharmacophore search and then prioritized using molecular docking and quantitative structure activity relationships.

Molecular fingerprint based similarity search analysis

Initially, a virtual library of ~4 million commercially available drug like compounds retrieved from the ZINC database was reduced by eliminating compounds distant from the three highly potent *M. tuberculosis* Type II dehydroquinase inhibitors 13, 28 and 31 (Table 1) on the basis of TGD and TGT 2D structural fingerprints. We computed the structural similarities using a 2-point (TGD) and a 3-point (TGT) pharmacophore-based fingerprint

using the Molecular Operating Environment (MOE) software [44]. Subsequently, we calculated the Tanimoto similarity coefficient (Tc) with compounds 13, 28 and 31. To ensure that each individual search yielded good number of compounds from the input database, we adjusted the Tc to 60% for the TGD and TGT fingerprint for compounds 13, 28 and 31. The hits retrieved using similarity searches were then merged and duplicates were removed that yielded a library of 268186 unique compounds.

3D pharmacophore search

To retrieve novel and diverse scaffolds for the identification of potential compounds targeting *M. tuberculosis* Type II dehydroquinase we subjected the 268186 compounds obtained in the 2D similarity analysis to a 3D pharmacophore search using the Unity module of Sybyl 7.1 [46]. Bound conformation of highest active inhibitor 3-nitrophenyl derivative of quinate (Compound 28) was used for the generation of pharmacophore hypothesis. In order to get putative bound conformation, we docked 3-nitrophenyl derivative of quinate (Compound 28) to binding pocket using FlexX. The purpose of molecular docking of highly potent inhibitors was to identify the structural features in context of their putative protein bound conformation as well as provide information for the development of 3D-pharmacophore model. Compound 28 adopted a similar conformation in the binding site and established similar contacts with the amino acid residues as the other crystal

Fig. 1 Scheme of the structure based virtual screening protocol for identification of M. tuberculosis Type II dehydroquinase targeted virtual library. Initially, ZINC database was subjected to 2D fingerprint similarity searches based on the highly active known inhibitors 13, 28 and 31 in Table 1. The threshold of the Tanimoto coefficient (Tc) for the search was fixed to 60% to select good number of compounds from the input database. In the next step, resulting database of compounds was subjected to 3D pharmacophore search and then to high-throughput docking. In the final step, selection of potent compounds was done based on receptor interaction fingerprint based scoring quantitative structure activity relationship model
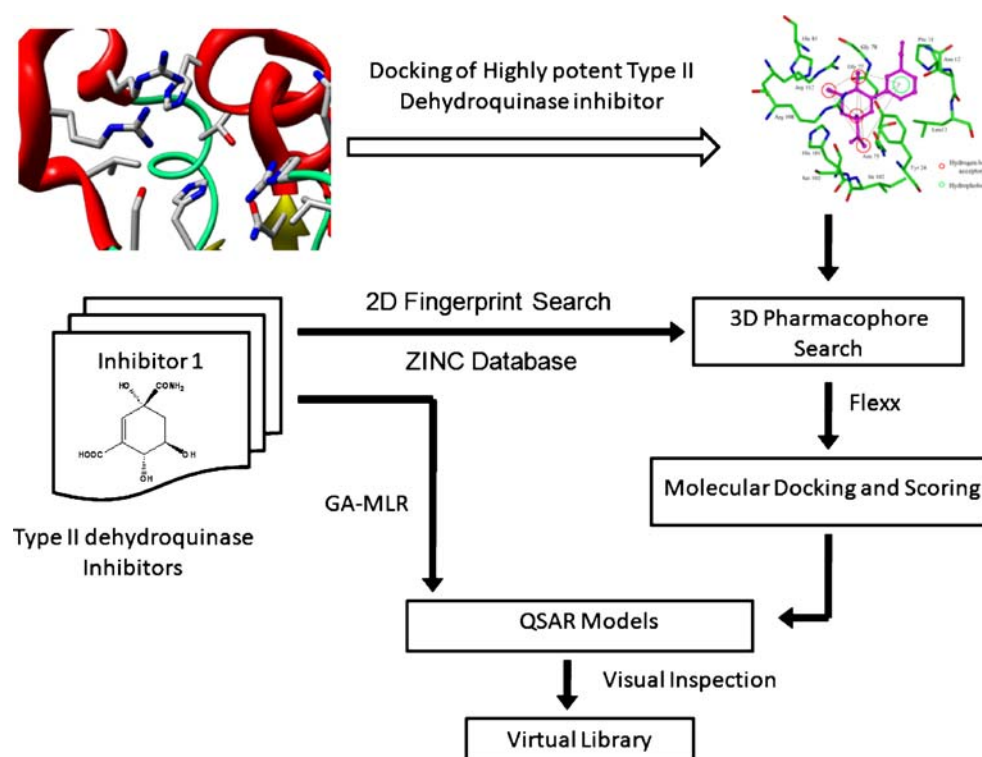
structure ligands 2,3-anhydro-quinic acid and 3-hydroxyimino quinic acid (PDB code 1H0R and 1H0S respectively) (Fig. 2). We defined the 3D pharmacophore on the basis of docked conformation using four hydrogen bond acceptor features and a hydrophobic feature. The hydrogen bond acceptor features represented in our hypothesis form hydrogen bonding interactions with Asn75, His81, Ile102, Ser103 and Arg112. The hydrophobic feature makes significant hydrophobic interactions with Pro11, Leu13 and Tyr24. The pharmacophore hypothesis was then used as a 3D structural query for further screening of 268186 compounds obtained in molecular fingerprint based similarity analysis using the Unity module of Sybyl 7.1 [46]. Pharmacophore based virtual screening yielded 4821 hits that met the specified requirements.

Molecular docking and scoring of the hits obtained from 3D pharmacophore search

All the 4821 hit compounds retrieved from the pharmacophore based screening were then subjected to molecular docking to the Type II dehydroquinase inhibitor binding site. *M. Tuberculosis* Type II dehydroquinase-FA6 binary complex (PDB entry code: 1H0R) 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 was generated. FlexX successfully docked 4808 out of 4821 hits subjected to molecular docking and 30 distinct poses for each 4808 compounds were further rescored using C-Score module of Sybyl7.1 to obtain G_Score [48], PMF_Score [49], D_Score [50] and Chem-Score [51] values so as to help in better hit prioritization. To

avoid any bias associated with individual scoring functions [67], the five scores from each ligand pose were normalized and combined to give an additional consensus score using CScore module of Sybyl7.1. Combined scoring functions perform in a superior fashion to the single scoring function and by their nature the combination of functions will ameliorate the effect of any particular unsuitable single function. The ligand poses with best consensus scores for each 4808 hits were then selected for further investigation.

A common problem in receptor based virtual screening is that some compounds are ranked well by scoring functions post-docking, although their respective pose is barely in the binding site leading to false Enrichments [68]. Second, and importantly, Warren et al. have recently observed, that from an assessment of 35 scoring functions, none were able to reliably identify the best-docked pose against a set of different targets [69]. Obviously, this leads to difficulty in prioritizing leads in a structure based virtual screen campaign against a particular target. To overcome this pitfall, we incorporated receptor knowledge based weighting approach along with conventional FlexX scoring in our virtual screening study to ensure that only realistic binders are prioritized. The knowledge based scoring scheme presented in this study is based on the incorporation of receptor−ligand interaction information from reference ligand to the receptor, *i.e.*, validated docked conformation of highest active compound, *i.e.*, 3-nitrophenyl derivative of quinic acid (compound **28**). Patterns of interactions with the receptor were modeled using binary ligand−receptor fingerprints (IFP) as described in the Section Receptor interaction fingerprints. Receptor interaction profile is used to filter out compounds with docked binding modes that are not similar to those observed in the docking predicted bound complex of Type II dehydroquinase with Compound **28**.

To ensure the ability of our scoring protocol where we incorporated receptor knowledge based weighting approach along with conventional FlexX scoring to discern true positives, enrichment studies were carried out on a test library consisting of known actives and decoys. Forty five known actives with experimentally reported K_i were collected from literature [19, 20, 38–40] and pooled with 500 decoys which were prepared in a fashion similar to the DUD decoy set [41]. As, compound compilation used in a test library can have profound effects on enrichment studies therefore compounds having comparable physicochemical properties to the known inhibitor were retrieved from the Maybridge screening library using Euclidean distance as similarity metric. This test library consisting of 545 compounds (45 actives and 500 decoys) was also subjected to molecular docking to Type II dehydroquinase binding site using the similar protocol. FlexX failed to find docking solution for 22 decoys which resulted in net test screening set of 523 compounds including 45 actives and 478 decoys.
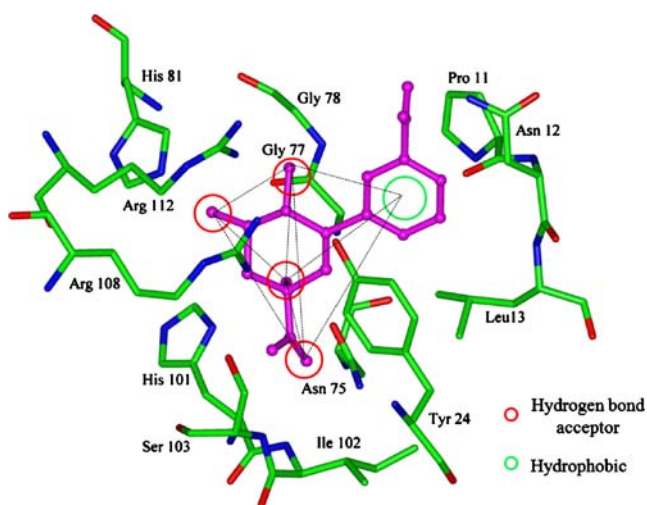

Fig. 2 3D pharmacophore query generated based on the putative bound conformation of Compound 28 (3-nitrophenyl derivative of quinic acid)
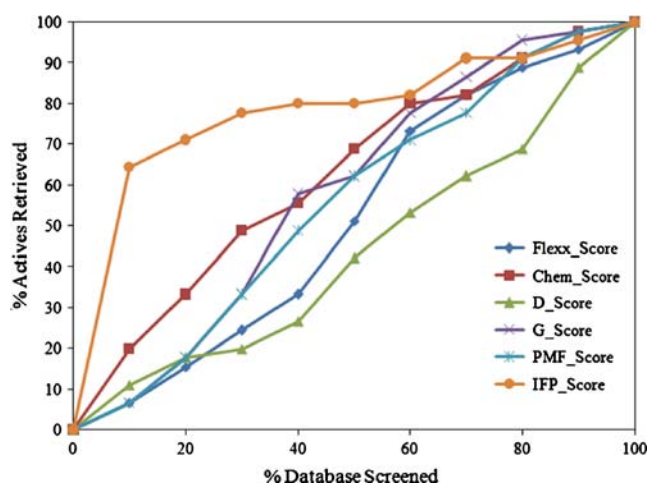
Fig. 3 Enrichment curves obtained by docking the screening set compounds consist of *M. tuberculosis* type II dehydroquinase actives and decoys into binding site (PDB entry code 1H0S) using FlexX. Selection and rank ordering of the docking solutions was performed using receptor interaction fingerprint based scoring scheme (IFP) and five conventional scoring functions (Flexx_Score, Chem_Score, D_Score, G_Score and PMF_Score)

As an indicator of performance, the enrichment factor was calculated as:

$$EF = \frac{\text{Known Binders}_{\text{sampled}}^{x\%}}{N_{\text{sampled}}^{x\%}} \times \frac{N_{\text{total}}}{\text{Known Binders}_{\text{total}}} \quad (3)$$

where Known Binders$_{\text{sampled}}^{x\%}$ is the number of known binders retrieved at $x\%$ of the database screened, $N_{\text{sampled}}^{x\%}$ is the number of compounds screened at $x\%$ of the database, known binders$_{\text{total}}$ is the number of known binders in the entire database and N_{total} is the number of compounds in the entire database. The enrichment plot (Fig. 3) shows a comparison of the performance of five conventional scoring function and receptor interaction fingerprint based scoring in retrieving the known actives among the top ranking compounds. It is clear that scoring functions behave quite differently, reflecting their diverse approaches to quantifying molecular interactions. As, the actual number of binders among the screening set was not known so the enrichment factors were calculated using 10, 20 and 50 % of the database screened and are presented in Table 2. Using the conventional scoring functions, the enrichment of known binders among the top ranked molecules was poor while the

receptor interaction fingerprint based scoring perform in superior fashion with enrichment factor values of 6.481, 3.576 and 1.609 for 10, 20 and 50% of the ranked database (Table 2 and Fig. 3). This implies that receptor interaction fingerprint based scoring is both effective and robust in its ability to recognize compounds that are known to interact with *M. tuberculosis* Type II dehydroquinase binding site and suggest that similarly high ranking compounds may display some affinity in vitro.

The design of new compounds should never be based on one single approach. Although receptor interaction fingerprint based scoring approach gave better result than the conventional scoring functions in post processing the docking results but better results can be achieved using the consensus of two approaches. A good approach to design a library targeting a particular protein might be to combine receptor interaction fingerprint based scoring with the conventional scoring functions by taking the top results identified separately by the two methods. The selection of compounds is illustrated in Fig. 4. As FlexX tries to determine the binding free energy, hits which have a good FlexX score (−20 to −35 kJ mol^{-1}) with CScore value of 5, *i.e.*, holds true for all five scoring functions (Flexx_score, G_Score, PMF_Score, D_Score and ChemScore) and good receptor interaction fingerprint score were selected. A Tanimoto similarity score (T_c) of 0.6 was taken as cutoff for the selection of compounds according to receptor interaction fingerprint based scoring. Fifty nine molecules were then selected for further investigation by using conventional scoring functions and by similarity searching the hits on the basis of reference inhibitor, using receptor interaction fingerprints and applying a T_c threshold value of 0.6.

Quantitative structure activity relationships models

The genetic algorithm multiple linear regression analysis was employed to derive the QSAR model explaining the inhibitory activities of Type II dehydroquinase inhibitors using training set of 35 compounds and test set of 10 compounds selected rationally using the diverse subset tool of MOE using MACCS keys and Tanimoto similarity coefficient. The application of the genetic algorithm for variable selection and subsequent development of multiple linear regression model using all the default parameters resulted in 13 descriptor model with non crossvalidated
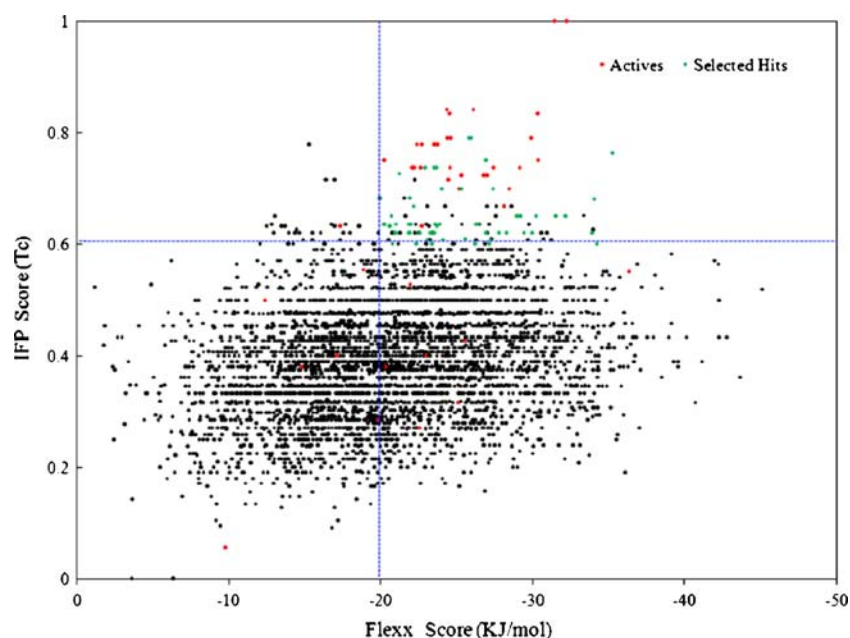
Table 2 Enrichment factors for the top-scored 10, 20 and 50 % of the ranked database

Scoring scheme						
	Flexx_Score	Chem_Score	D_Score	G_Score	PMF_Score	IFP
10%	0.670	2.011	1.117	0.670	0.670	6.481
20%	0.782	1.676	0.894	0.894	0.894	3.576
50%	1.028	1.385	0.849	1.251	1.251	1.609

Fig. 4 Receptor interaction fingerprint-based compound selection for potential type II dehydroquinase binding compounds with compound **28** (3-nitropheny derivative of quinic acid) used as reference ligand for compound selection. A Tanimoto similarity score of 0.6 and Flexx_-Score of −20.00 kJ mol^{-1} taken as cutoff for the selection of compounds. Black points represents total virtual screening hits retrieved after 2D similarity and 3D pharmacophore search while red and green points represents known actives and selected hits respectively

correlation coefficient (r^2) of 0.952 and Leave-one-out cross validated correlation coefficient (q^2) of 0.885. This model also explained well the variance in the activities of the test set with predictive correlation coefficient (r_{pred}^2) of 0.982 and had good statistical significance (F-value=33.814). The regression equation and the statistical items were as follows:

$$20.7338 + (-2.43341 \times BCUT_{SLOGP_1}) + (-0.228062 \times DLI(10)) + (-21.3736 \\ \times GCUT_{SMR_2}) + (0.0143473 \times PEOE_{VSA_{POL}}) + (-0.0774612 \times PEOE_{VSA_{PPOS}}) \\ + (0.875022 \times PM3dipole) + (-3.11956 \times Q_{VSA_{FPOS}}) + (0.0355627 \\ \times SMR_{VSA6}) + (-0.139282 \times bpol) + (-2.86432 \times dipole) + (-8.6512 \\ \times glob) + (-0.268099 \times kC_{ssssC}) + (-2.66804 \times petitjeanSC)$$

(4)

$$r^2 = 0.952 \ rmse = 0.2037 \ q^2 = 0.885 \ LOF$$
$$= 0.6275 \ F = 33.814 \ r_{pred}^2 = 0.982.$$

(5)

The predicted activities and correlation between the predicted activities and the experimental activities of the training and test set compounds are depicted in Table 1 and Fig. 5.

Selection and binding mode of potential hits

Ligand based methods of analysis such as 3D-QSAR 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 GA-MLR 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 GA-MLR model to virtual screening hits identified using 2D similarity, 3D pharmacophore and molecular docking incorporating additional knowledge from receptor inter-action fingerprints for the prediction of biological
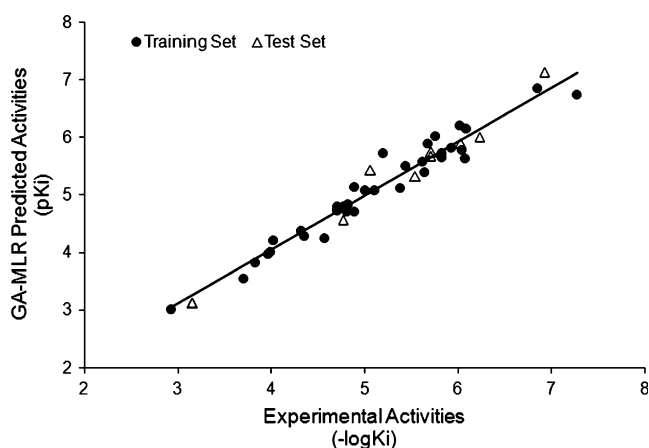

Fig. 5 Correlation between experimental and GA-MLR predicted activities of *M. tuberculosis* type II dehydroquinase inhibitors

Table 3 Chemical structures of putative *M.tuberculosis* TypeII dehydroquinase binders and their respective docking scores, predicted activity, molecular weight and LogP identified using virtual screening and QSAR

S.No.	ZincID	Structure	Flexx Score	Mol. Wt	LogP	pKi
1.	ZINC06677576		-23.64	419.48	0.86	11.125
2.	ZINC03552284		-21.88	282.25	3.39	10.734
3.	ZINC09318451		-20.62	489.57	1.00	10.619
4.	ZINC05577070		-20.01	378.45	5.67	9.947
5.	ZINC04484679		-27.16	355.44	0.74	9.917
6.	ZINC03264172		-21.96	290.32	1.20	9.666
7.	ZINC06765153		-24.06	415.87	3.27	9.472
8.	ZINC09490981		-25.56	472.59	2.97	9.292
9.	ZINC01786822		-22.97	390.86	2.36	9.111

Table 3 (continued)
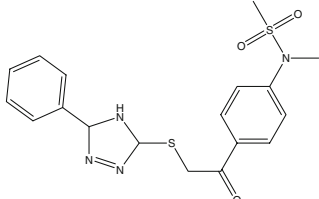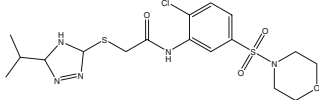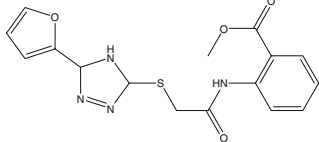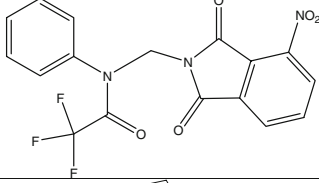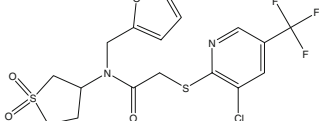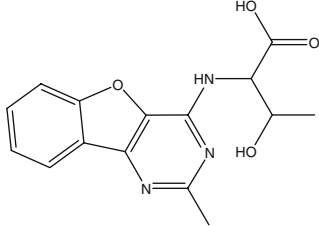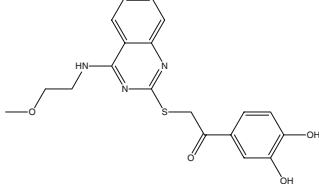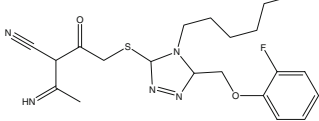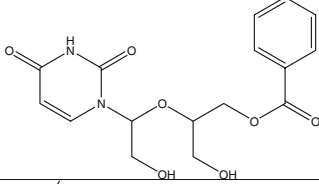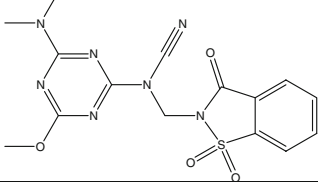
10.	ZINC05262611		-23.09	402.50	2.62	9.107
11.	ZINC09514151		-22.60	459.98	1.43	8.479
12.	ZINC02853073		-26.44	358.38	1.97	8.218
13.	ZINC00652050		-34.20	395.29	1.98	8.138
14.	ZINC09173622		-23.66	472.94	2.50	7.958
15.	ZINC05215425		-20.74	301.30	1.22	7.911
16.	ZINC02948647		-26.97	385.44	2.65	7.786
17.	ZINC09630683		-22.13	431.54	3.88	7.719
18.	ZINC03624369		-20.30	350.33	-0.74	7.706
19.	ZINC00857312		-27.37	389.40	-0.49	7.638

Table 3 (continued)
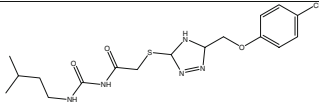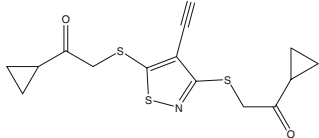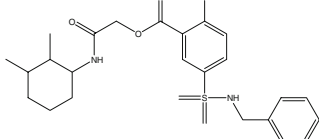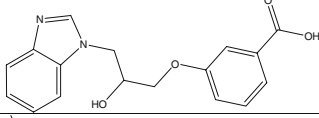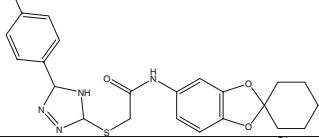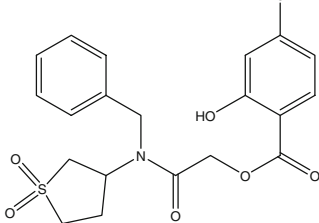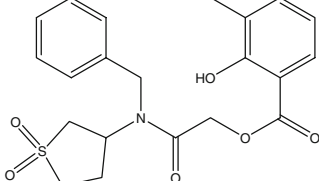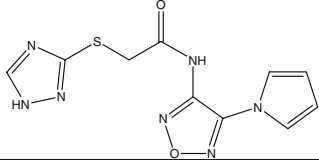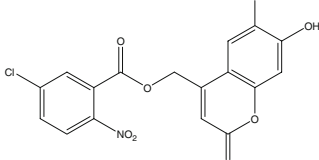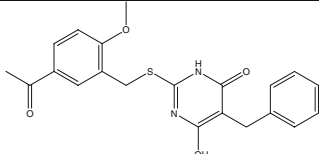
20.	ZINC03393572		-26.26	411.91	3.41	7.560
21.	ZINC04871579		-23.55	338.48	2.57	7.527
22.	ZINC09539647		-26.37	474.58	4.28	7.456
23.	ZINC04905087		-26.20	312.32	2.39	7.426
24.	ZINC09514369		-30.00	436.54	5.86	7.377
25.	ZINC09144730		-25.76	437.90	2.65	7.257
26.	ZINC04923168		-25.91	417.48	2.32	7.118
27.	ZINC01422371		-22.04	291.29	0.94	7.014
28.	ZINC03348013		-33.94	393.73	2.79	6.920
29.	ZINC04148781		-20.98	396.47	4.12	6.873

Table 3 (continued)
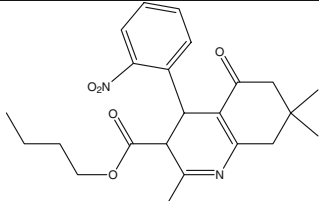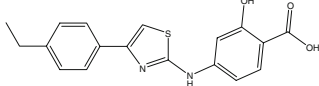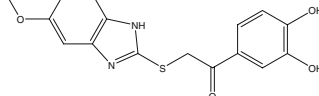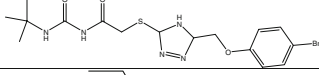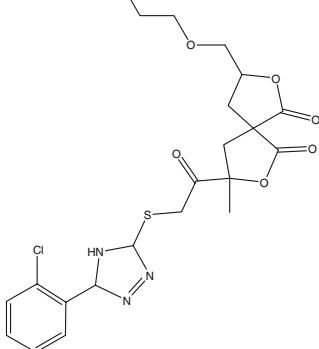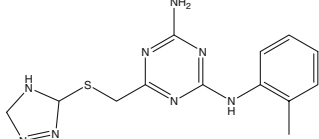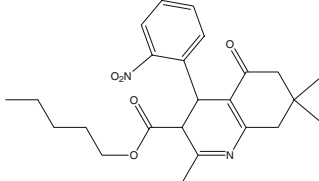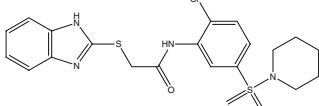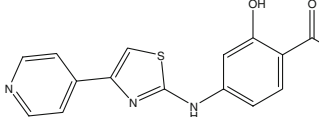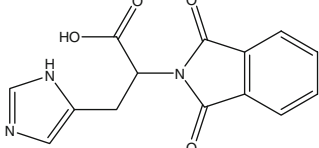
30.	ZINC04059796		-21.86	414.50	2.89	6.858
31.	ZINC00070699		-29.02	340.40	4.37	6.798
32.	ZINC09156838		-23.71	330.36	3.00	6.645
33.	ZINC02616424		-23.54	442.34	3.07	6.622
34.	ZINC05225957		-34.08	508.00	4.80	6.515
35.	ZINC05620766		-22.41	314.38	0.67	6.505
36.	ZINC04059683		-21.00	428.53	3.34	6.500
37.	ZINC09373218		-24.14	465.00	3.38	6.372
38.	ZINC00612918		-29.21	313.34	2.36	6.344
39.	ZINC00129993		-28.23	285.26	0.27	6.204

Table 3 (continued)

40.	ZINC03517387		-23.39	298.37	0.96	6.129
41.	ZINC01141392		-30.21	357.39	3.06	6.093
42.	ZINC04494538		-21.97	361.39	2.12	6.034
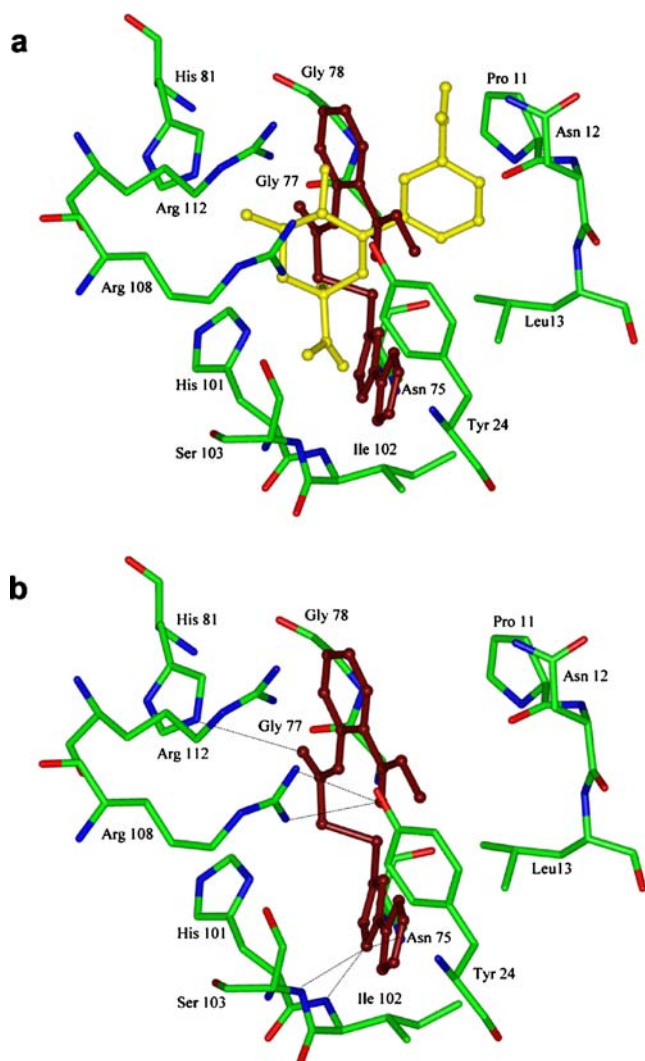
Fig. 6 a & b. Binding mode of ZINC02853073 (Brown color) shown with Compound 28 (Yellow). Dashed lines indicate hydrogen bond between ZINC02853073 and protein

activity. Thus based on the findings derived from the developed GA-MLR model, the biological activity of 59 selected hits was predicted. As expected, the predicted activity of almost half of the virtual screening hits, *i.e.*, 47.4% was higher than the reported experimental activity of highest active Type II dehydroquinase inhibitor (-logK$_i$ of 7.27 for 3-nitrophenyl derivative of quinic acid, (compound **28** in Table 1)). The predicted activity of 30.5% of the virtual screening hits was in moderate range, *i.e.*, 7.27< pk$_i$ >6.00. The virtual screening hits with predicted activity in higher and moderate range were then subjected to visual inspection for final selection of molecular starting points for the design and identification of Type II dehydroquinase inhibitors.

Recent molecular modeling studies [19, 20, 38] and our molecular docking studies on Type II dehydroquinase inhibitors into the active site of *M. tuberculosis* Type II dehydroquinase which is composed of Pro11, Asn12, Leu13, Arg19, Tyr24, Asn75, Gly77, Gly78, His81, His101, Ile102, Ser103, Arg108 and Arg112 revealed the inhibitors are held in the active site with their quinate core in a similar position as that adopted by crystal structure ligands 2,3-anhydro-quinic acid and 3-hydroxyimino quinic acid (PDB code 1H0R and 1H0S respectively). C1 ring carboxylate is held in place by hydrogen bonding interaction with two backbone amides from Ile102 and Ser103 whereas C1 hydroxyl group forms hydrogen bonding interaction with His101 and Asn75. Molecular docking predicted binding puts aromatic ring emerging from C3 substitutions of quinate core of highly active inhibitors into the pocket formed by Arg108 and Arg112. Another key binding interactions observed in case of *Streptomyces coelicolor* type II dehydroquinase were edge to face stacking interactions of aromatic ring with Tyr28 and cation

π interaction with Arg23 (Tyr24 and Arg19 in case of *M. tuberculosis* type II dehydroquinase). However we were unable to get clear cut idea for these interactions due to lack of these important residues in crystal structure (Tyr24 in 1H0R and Arg19 in 1H0S).

The virtual screening hits with predicted activity in higher and moderate range were then inspected visually considering the above mentioned interactions for *M. Tuberculosis* type II dehydroquinase inhibitor binding and a set of 42 molecules was finally selected. The structures of the selected compounds are shown in Table 3, together with their corresponding Flexx_Score, GA-MLR predicted activity, molecular weight and calculated log P values. The virtual screening process derived from this integrated computational approach allowed the identification of several chemotypes, not previously reported as Type II dehydroquinase inhibitors, which includes sulfanyl triazoles, sulfanyl, phenyl and pyridine thiazoles, sulfanyl quinazoline, pyrimidines, pyridine and triazine derivatives *etc*. The promising compounds share common features like a nitrogen rich scaffold triazoles, thiazoles, quinazolines, pyrimidine, pyridine, triazines *etc*.) and various hydrophobic group connected to the scaffold. Among these sulfanyl triazoles scaffold forms the major class and could be exploited for further optimization. The predicted binding mode of a sulfanyl triazole compound **ZINC02853073** displaying high docking score and QSAR predicted activity is described in Fig. 6 along with the putative bound conformation of compound **28** with *M. tuberculosis* type II dehydroquinase. Like the bound conformation of compound **28** shown in yellow color in Fig. 6a, **ZINC02853073** is anchored to the cavity by combination of hydrogen bonding interactions with Asn75, His81, Ile102, Ser103 and Arg108. Triazole core nitrogen interacts *via* hydrogen bonds with the backbone amides of Ile102 and Ser103 and with side chain nitrogens of Asn75. Binding of **ZINC02853073** is also stabilized by hydrogen bonding interaction between sulfanyl oxygen and sidechain nitrogen of His81 and between benzoic acid ester oxygen and sidechain nitrogen of Arg108 as shown in Fig. 6b.

Conclusions

Virtual screening of small molecule libraries has become a standard practice in drug discovery. Here, in this study, we emphasize the need to integrate ligand and structure-based drug design approaches to maximize the benefits for individual targets. The multistep strategy combines a ligand-based (2D similarity search, 3D pharmacophore) and structure based (molecular docking and interaction fingerprint based scoring) virtual screening for building an enriched library of small molecules with a QSAR for screening against the *M. tuberculosis* Type II dehydroquinase. The important feature of this computational approach is the enhancement of its predictive power by incorporating an exhaustive consensus scoring process using a rank-by-rank strategy in order to provide a list of suitable compounds to be considered for bioscreening. On the basis of 2D fingerprints, a similarity search on the ZINC database was performed using highly potent inhibitors of M. tuberculosis Type II dehydroquinase as reference structures. After 3D pharmacophore filtering, the resulting enriched library was docked into the binding site to provide a list of putative ligands ranked according to consensus scoring using conventional Flexx scoring functions and receptor interaction fingerprints (IFP). Additionally, known active compounds were used for the development of ligand based QSAR model in order to use them to predict activity of hitherto unsynthesized molecules. The resulting ligand and structure-based virtual screening allowed identification several scaffolds like sulfanyl triazoles, sulfanyl, phenyl and pyridine thiazoles, sulfanyl quinazoline, pyrimidines, pyridine and triazine derivatives *etc*. that have not been previously characterized in the scientific literature as *M. tuberculosis* type II dehydroquinase inhibitors and can be useful to a medicinal chemist or combinatorial chemist to pick up the new molecular starting points for medicinal chemistry optimization for the design of novel M. tuberculosis type II dehydroquinase inhibitors.

References

1. Abell C (1999) In: Sankawa U (ed) Comprehensive natural products chemistry, Vol 1. Pergamon, Elsevier, Oxford
2. Haslam E (1993) Shikimic acid: metabolism and metabolites. Wiley, Chichester, UK
3. Bentley R (1990) The Shikimate pathway - a metabolic tree with many branches. Crit Rev Biochem Mol Biol 25:307–384
4. Gourley DG, Shrive AK, Polikarpov I, Krell T, Coggins JR, Hawkins AR, Isaacs NW, Sawyer L (1999) The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. Nat Struct Biol 6:521–525
5. Butler JR, Alworth WL, Nugent MJ (1974) Mechanism of dehydroquinase catalysed dehydration. I. Formation of a Schiff base intermediate. J Am Chem Soc 96:1617–1618
6. Shneier A, Kleanthous C, Deka R, Coggins JR, Abell C (1991) Observation of an imine intermediate on dehydroquinase by electrospray mass spectrometry. J Am Chem Soc 113:9416–9418
7. Harris JM, Kleanthous C, Coggins JR, Hawkins AR, Abell C (1993) Different mechanistic and stereochemical courses for the reactions catalyzed by type-I and type-II dehydroquinases. J Chem Soc Chem Commun 13:1080–1081

8. Shneier A, Harris JM, Kleanthous C, Coggins JR, Hawkins AR, Abell C (1993) Evidence for opposite stereochemical courses for the reactions catalyzed by type-I and type-II dehydroquinases. Bioorg Med Chem Lett 3:1399–1402

9. Harris JM, Gonzalez-Bello C, Kleanthous C, Hawkins AR, Coggins JR, Abell C (1996) Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases. Biochem J 319:333–336

10. Garbe T, Servos S, Hawkins A, Dimitriadis G, Young D, Dougan G, Charles I (1991) The Mycobacterium tuberculosis shikimate pathway genes: evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases. Mol Gen Genet 228:385–392

11. Frederickson M, Parker EJ, Hawkins AR, Cogins JR, Abell C (1999) Selective inhibition of type II dehydroquinase. J Org Chem 64:2612–2613

12. Frederickson M, Coggins JR, Abell C (2002) Vinyl fluoride as an isoelectronic replacement for an enolate anion: Inhibition of type II dehydroquinases. Chem Commun 17:1886–1887

13. Le Sann C, Abell C, Abell AD (2002) A simple method for the preparation of 3-hydroxyiminodehydroquinate, a potent inhibitor of type II dehydroquinase. J Chem Soc Perkin Trans 18:2065–2068

14. Frederickson M, Roszak AW, Coggins JR, Lapthorn AJ, Abell C (2004) (1R, 4 S, 5R)-3-Fluoro-1, 4, 5-trihydroxy- 2-cyclohexene-1-carboxylic acid: The fluoro analogue of the enolate intermediate in the reaction catalyzed by type II dehydroquinases. Org Biomol Chem 2:1592–1596

15. Roszak AW, Robinson DA, Krell T, Hunter IS, Frederickson M, Abell C, Coggins JR, Lapthorn AJ (2002) The structure and mechanism of the type II dehydroquinase from Streptomyces coelicolor. Structure 10:493–503

16. Toscano MD, Frederickson M, Evans DP, Coggins JR, Abell C, Gonzalez-Bello C (2003) Design, synthesis and evaluation of bifunctional inhibitors of type II dehydroquinase. Org Biomol Chem 1:2075–2083

17. Gonzalez-Bello C, Lence E, Toscano MD, Castedo L, Coggins JR, Abell C (2003) Parallel solid-phase synthesis and evaluation of inhibitors of Streptomyces coelicolor type II dehydroquinase. J Med Chem 46:5735–5744

18. Toscano MD, Stewart KA, Coggins JR, Lapthorn AJ, Abell C (2005) Rational design of new bifunctional inhibitors of type II dehydroquinase. Org Biomol Chem 3:3102–3104

19. Prazeres VF, Sánchez-Sixto C, Castedo L, Lamb H, Hawkins AR, Riboldi-Tunnicliffe A, Coggins JR, Lapthorn AJ, González-Bello C (2007) Nanomolar competitive inhibitors of Mycobacterium tuberculosis and Streptomyces coelicolor type II dehydroquinase. ChemMedChem 2:194–207

20. Sánchez-Sixto C, Prazeres VF, Castedo L, Lamb H, Hawkins AR, González-Bello C (2005) Structure-based design, synthesis, and biological evaluation of inhibitors of Mycobacterium tuberculosis type II dehydroquinase. J Med Chem 48:4871–4881

21. Bleicher KH, Böhm H, Müller K, Alanine AI (2003) Hit and lead generation: beyond high-throughput screening. Nat Rev Drug Discovery 2:369–378

22. Bohacek RS, McMartin C, Guida WC (1996) The art and practice of structure-based drug design: a molecular modeling perspective. Med Res Rev 16:3–50

23. Lipinski C, Hopkins A (2004) Navigating chemical space for biology and medicine. Nature 432:855–861

24. Walters WP, Stahl MT, Murcko MA (1998) Virtual screening - an overview. Drug Discovery Today 3:160–178

25. Eglen RM, Schneider G, Böhm H (2000) In: Böhm H, Schneider G (eds) Virtual screening for bioactive molecules. Wiley-VCH, Weinheim

26. Drewry DH, Young SS (1999) Approaches to the design of combinatorial libraries. Chemom Intell Lab Syst 48:1–20

27. Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discovery 1:882–894

28. Miller JL (2006) Recent developments in focused library design: targeting gene-families. Curr Top Med Chem 6:19–29

29. Lahana R (1999) How many leads from HTS. Drug Discovery Today 4:447–448

30. Willett P (2006) Similarity-based virtual screening using 2D fingerprints. Drug Discovery Today 11:1046–1053

31. Mason JS, Good AC, Martin EJ (2001) 3-D pharmacophores in drug discovery. Curr Pharm Des 7:567–597

32. Schneider G, Böhm H (2002) Virtual screening and fast automated docking methods. Drug Discovery Today 7:64–70

33. Winkler DA (2002) The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery. Brief Bioinf 3:73–86

34. Schneider G, Fechner U (2005) Computer-based de novo design of drug-like molecules. Nat Rev Drug Discovery 4:649–663

35. Gozalbes R, Simon L, Froloff N, Sartori E, Monteils C, Baudelle R (2008) Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. J Med Chem 51:3124–3132

36. Kumar A, Chaturvedi V, Bhatnagar S, Sinha S, Siddiqi MI (2008) Knowledge based identification of potent antitubercular compounds using structure based virtual screening and structure interaction fingerprints. J Chem Inf Model 49:35–42

37. Kumar A, Siddiqi MI (2009) Virtual screening against Mycobacterium tuberculosis dihydrofolate reductase: suggested workflow for compound prioritization using structure interaction fingerprints. J Mol Graph Model 27:476–488

38. Payne RJ, Peyrot F, Kerbarh O, Abell AD, Abell C (2007) Rational design, synthesis, and evaluation of nanomolar type II dehydroquinase Inhibitors. Chem Med Chem 2:1015–1029

39. Toscano MD, Payne RJ, Chiba A, Kerbarh O, Abell C (2007) Nanomolar inhibition of type II dehydroquinase based on the enolate reaction mechanism. Chem Med Chem 2:101–112

40. González-Bello C, Castedo L (2007) Progress in type II dehydroquinase inhibitors: from concept to practice. Med Res Rev 27:177–208

41. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49:6789–6801

42. Venhorst J, Núñez S, Terpstra JW, Kruse CG (2008) Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. J Med Chem 51:3222–3229

43. Evers A, Hessler G, Matter H, Klabunde T (2005) Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. J Med Chem 48:5448–5465

44. MOE, Version 2007 Chemical Computing Group, Inc. Montreal, Quebec, Canada

45. Rogers DJ, Tanimoto TT (1960) A computer program for classifying plants. Science 132:1115–1118

46. SYBYL Molecular Modeling System. [7.1] (2006) TRIPOS, Assoc, Inc. St-Louis, MO,

47. Hurst T (1994) Flexible 3D searching - the directed tweak technique. J Chem Inf Comput Sci 34:190–196

48. Jones G, Willett P, Glen G (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol 245:43–53

49. Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 42:791–804

50. Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comput Chem 13:505–524

51. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 11:425–445

52. Marcou G, Rognan D (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 47:195–207

53. Kier LB, Hall LH (1976) Molecular connectivity in chemistry and drug research. Academic Press, New York

54. Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. Wiley, New York

55. Randi M (1975) On characterization on molecular branching. J Am Chem Soc 97:6609–6615

56. Kier LB (1985) A shape index from molecular graphs. Quant Struct-Act Relat 4:109–116

57. Kier LB (1987) Inclusion of symmetry as a shape attribute in kappa-index analysis. Quant Struct-Act Relat 6:8–12

58. Petitjean M (1992) Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds. J Chem Inf Comput Sci 32:331–337

59. Balaban AT (1979) Five new topological indices for the branching of tree-like graphs. Theor Chim Acta 53:355–375

60. Balaban AT (1982) Highly discriminating distance-based topological index. Chem Phys Lett 89:39–404

61. Wiener H (1947) Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons. J Am Chem Soc 69:2636–2638

62. Wiener H (1947) Structural determination of paraffin boiling points. J Am Chem Soc 69:17–20

63. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. Tetrahedron 36:3219–3228

64. Stanton D, Jurs P (1990) Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. Anal Chem 62:2323–2329

65. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure activity relationships and quantitative structure property relationships. J Chem Inf Comput Sci 34:854–866

66. Irwin JJ, Shoichet BK (2005) ZINC - a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182

67. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB (2002) Consensus scoring for ligand/protein interactions. J Mol Graph Model 20:281–295

68. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49:5851–5855

69. Warren GL, Andrews CW, Capelli AM, Clarke B, Lalonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931