



# Virtual screening against *Mycobacterium tuberculosis* dihydrofolate reductase: Suggested workflow for compound prioritization using structure interaction fingerprints

Ashutosh Kumar, Mohammad Imran Siddiqi \*

Molecular and Structural Biology Division, Central Drug Research Institute, Lucknow 226 001, India

## ARTICLE INFO

### Article history:

Received 12 June 2008

Received in revised form 18 August 2008

Accepted 19 August 2008

Available online 28 August 2008

### Keywords:

Pharmacophore hypothesis

Virtual screening

Molecular docking

Structure interaction fingerprints

## ABSTRACT

In this study, we suggest a new workflow for the identification and prioritization of potential compounds targeted against *Mycobacterium tuberculosis* dihydrofolate reductase, an important folate cycle enzyme and a validated target for the development of anti-tubercular agents. First, we have performed an integrated pharmacophore and structure-based virtual screening using Maybridge small molecule database, subsequently interaction patterns from known actives to the receptor were applied for scoring and ranking the virtual screening hits using structure interaction fingerprint (SIFT)-based similarity approach. In addition, agglomerative hierarchical clustering of the structure interaction fingerprints permits the easy separation of active from inactive binding modes. Using this approach we screened 59275 Maybridge compounds and 20 compounds were prioritized as promising virtual screening hits. Though using a receptor interaction scoring approach, the results were not biased toward the chemical classes of the known actives and the proposed compounds were structurally diverse with low molecular weights and structural complexities. Our results suggest that structure-based virtual screening coupled with the SIFT should be a valuable tool for prioritization of virtual screening hits.

© 2008 Elsevier Inc. All rights reserved.

## 1. Introduction

The ongoing selection of multidrug-resistant strains of *Mycobacterium tuberculosis* has markedly reduced the effectiveness of the standard treatment regimens. Thus, there is an urgent need for new drugs that are potent inhibitors of *M. tuberculosis* that exhibit favorable resistance profiles and that are well tolerated by patients. One promising drug target for treatment of mycobacterial infections is dihydrofolate reductase a key enzyme in the folate cycle [1,2] which supplies one-carbon units, derived from the action of serine hydroxymethyltransferase [3,4] on L-serine, for the biosynthesis of deoxythymidine monophosphate (dTMP). Inhibition of the folate cycle leads to interruption of the supply of thymidine and thus to inhibition of DNA biosynthesis and inhibition of proliferation of cells. Inhibition of proliferation is a useful goal in the therapy of bacterial and protozoal infections [5]. Although DHFR does not represent a new target, there is still enthusiasm for the development of DHFR inhibitors, particularly with regard to mycobacteria [6–10]. A unique feature of DHFR is the selectivity that is possible in the design of inhibitor. This makes

it an ideal “old” target for rational and effective drug design for anti-mycobacterial agents.

The drug discovery process is notoriously time-consuming and expensive [11] yet new drugs are required as it remains numerous unmet clinical needs in many disease indications. The number of potential target 3D structures is increasing at the Protein Data Bank (PDB) [12] and the number of drug/lead-like compounds is estimated to be at least  $10^{24}$  [13]. Thus, to deal with such a large amount of data and to facilitate the drug discovery process, *in silico* virtual screening and computer-aided drug design have become increasingly important in drug discovery [14]. How to intelligently leverage the 3D structural information of target molecules and use it in designing target-focused leads is of great interest in the field. Some techniques have been available for leveraging the structural information of the target molecules into library design and filtering and one of such approach is the 3D pharmacophore model. If a collection of known active molecules is available, 3D pharmacophore models can be generated from these compounds by extracting the common spatial arrangement of pharmacophoric features. These models can be applied to filter a large library and identify compounds that also satisfy the pharmacophore. If there is an experimentally determined high-resolution 3D structure of the target available, structure-based drug design can be performed which is associated with docking. Recent studies have demonstrated

\* Corresponding author.

E-mail addresses: [imsiddiqi@yahoo.com](mailto:imsiddiqi@yahoo.com), [mi\\_siddiqi@cdri.res.in](mailto:mi_siddiqi@cdri.res.in) (M.I. Siddiqi).

that docking algorithms and the underlying energy-scoring functions are usually able to reproduce the crystallographic binding modes; however, these same scoring functions are poor at identifying the correct binding mode for the virtual screening hits, leading to a large number of high-scoring false positives that result in lower enrichment rates for virtual screening [15].

Structural interaction fingerprint (SIFt), a novel method for efficiently representing, visualizing, and analyzing massive amounts of structures. SIFt has been proven to be useful in facilitating post-docking analysis, virtual screening, and database mining of structural data [16–19]. Key to the SIFt method is the generation of structural interaction fingerprints, 1D binary bit-string that represent important target–ligand binding interactions, thus making the structures amenable to easy mathematical manipulation and comparison. Recent studies have shown that by combining SIFt and other conventional scoring functions, we can achieve much better confidence in reproducing the true binding modes of the compounds and thereby obtain improved library enrichments from virtual screening [16–19].

In continuation of our efforts to discover new chemical entities endowed with activity against *M. tuberculosis*, we have employed an integrated database screening strategy involving three approaches: pharmacophore hypothesis based 3D database search, protein structure-based molecular docking and similarity based screening using structure interaction fingerprints.

## 2. Computational methods

### 2.1. Dataset

A total of 78 inhibitors of *MtDHFR* were collected from the PDB (ligands complexed with protein) and various literature sources [6–10]. The compounds were built using Builder module of InsightII 2000.1 and then optimized using CVFF force field. For the validation of the pharmacophore models, a test set database was prepared consisting of these 78 known inhibitors of *MtDHFR* spiked into a subset of World of Molecular Bioactivity (WOMBAT) database consisting of 1207 molecules, active against different proteins other than used in present study, considered here as inactive.

### 2.2. Pharmacophore hypothesis generation and virtual screening

A set of five pharmacophore hypotheses were generated based on three available *MtDHFR*–inhibitor complexes taken from the PDB (PDBid: 1DF7, 1DG5 and 1DG7). Pharmacophore hypothesis generation was performed with the UNITY and GASP software available with SYBYL7.1 [20]. Virtual screening was carried out using these pharmacophore hypotheses using UNITY software. Maybridge small molecule database was used to screen the compounds. The flexible search methodology was employed as it takes into account the conformational flexibility of molecules.

### 2.3. Docking and scoring

The compounds obtained from the pharmacophore-based virtual screening and also the known inhibitors of *MtDHFR* were docked into the binding pocket of *MtDHFR* using the FlexX program interfaced with SYBYL7.1 [20]. FlexX employs a fast algorithm for flexible docking of small ligands into a fixed protein-binding site using an incremental construction process [21]. Standard parameters of the FlexX program as implemented in SYBYL7.1 were used during docking. To further evaluate the docking experiment, the G\_Score [22], PMF\_Score [23], D\_Score [24] and ChemScore [25] values were estimated using the CScore

module of SYBYL7.1 [20]. 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.

### 2.4. Construction of structure interaction fingerprints (SIFt)

SIFt stands for structural interaction fingerprint, which is a 1D binary fingerprint representation of the intermolecular interactions in a 3D protein–inhibitor complex. The SIFt protocol and implementation reported previously [16–19] were used to generate the interaction fingerprints. The SIFt is generated by first defining the union of those residues that are in contact between the protein and the small molecule complex. A uniform list of amino acid residues involved in ligand binding was used for calculating the SIFt for all the ligands. In addition, we identified the protein atoms that were involved in hydrogen-bonding interactions with the ligands, using the program HBPLUS [26] with default settings. The program calculated and listed all possible hydrogen bond donor and acceptor pairs in the structure that satisfy predefined geometric criteria. The hydrogen-bonding pairs between protein and ligand were extracted for subsequent analysis. After all the ligand binding site residues are identified and all the protein–ligand intermolecular interactions are calculated, the next step was to classify these interactions. Current implementation of SIFt uses seven bits for each binding site residue, representing seven different types of interactions. The seven bits are switched on or off if the following interactions are observed: (1) if a contact is involved at this position; (2) whether it involves the main-chain atoms; (3) whether it involves the side chain; (4) whether it is a polar interaction; (5) whether it is a non-polar interaction; (6) whether it is a hydrogen bond acceptor; (7) whether it is a hydrogen bond donor.

By doing so, each residue is represented by a seven-bit-long bit string. The whole interaction fingerprint of the complex is finally constructed by sequentially concatenating the bit string of each binding site residue together, according to the ascendant residue number order. This results in each interaction fingerprint being the same length, enabling easy comparison of interactions at a particular binding site position across a series of complexes.

### 2.5. Similarity analysis and hierarchical clustering of structure interaction fingerprints

We have used the Rogers and Tanimoto coefficient (RTc), commonly used for binary data [27] as the quantitative measure of bit string similarity. The RTc between two bit strings *A* and *B* is defined as:

$$RTc(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

where  $|A \cap B|$  is the number of ON bits common in both *A* and *B* and  $|A \cup B|$  is the number of ON bits present in either *A* or *B*. The structure interaction fingerprint represents the binding mode of a ligand to a target protein; similar fingerprints imply that the corresponding ligands make similar interactions with the protein. We applied a hierarchical clustering methodology to analyze the fingerprints for each test case. Interaction fingerprints were clustered by using an agglomerative hierarchical clustering approach [28], applying the Rogers and Tanimoto coefficient as similarity measurements. SYSTAT 12 [29] was used for hierarchical clustering of SIFt. Clusters of protein–ligand complex structures were manually selected based on the dendrogram of their interaction fingerprints.

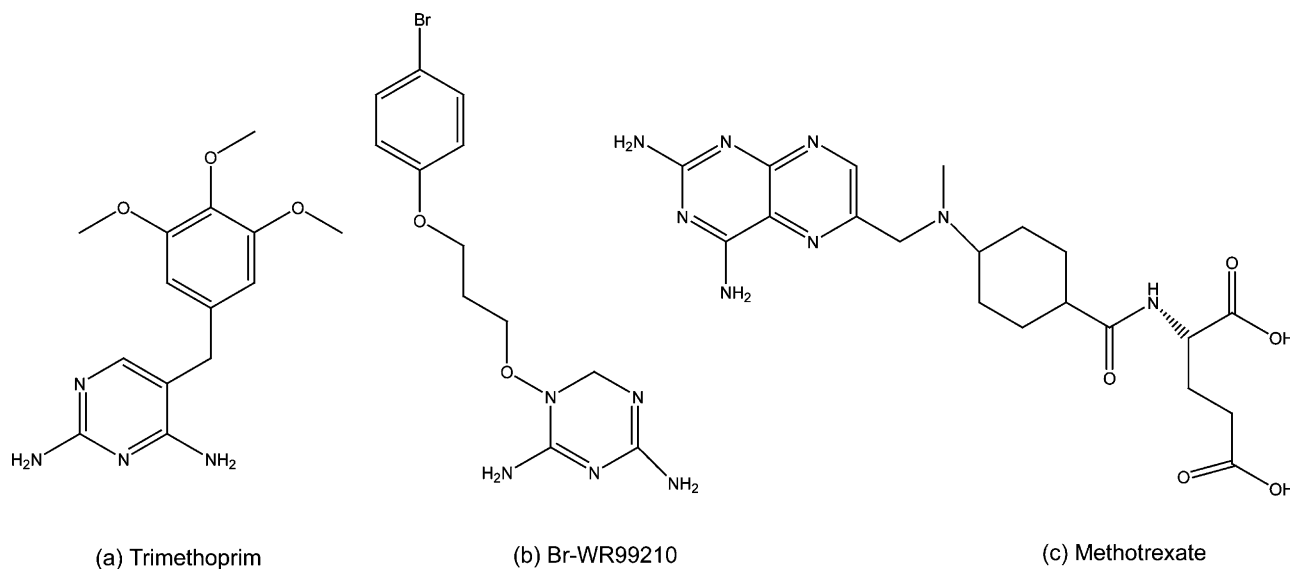


Fig. 1. MtDHFR inhibitors selected for pharmacophore hypothesis generation.

## 2.6. Hardware and software

InsightII 2000.1 [30] and Sybyl 7.1 [20] were used for molecular modeling on a SGI Origin 300 workstation equipped with  $4 \times 600$  Mhz R12000 processors.

## 3. Result and discussion

In this study, we have employed an integrated approach to identify potential MtDHFR hits using ligand and structure-based virtual screening, subsequently followed by structure interaction

fingerprints generation to prioritize the virtual screening compounds.

### 3.1. Pharmacophore hypothesis generation

A set of five pharmacophore hypotheses were generated using three available MtDHFR PDB complexed with inhibitors (PDBid: 1DG5, 1DG7 and 1DF7) [7]. The inhibitors used for pharmacophore generation are listed in Fig. 1 and five pharmacophore hypothesis generated from the PDB complexed ligands are presented in Fig. 2.

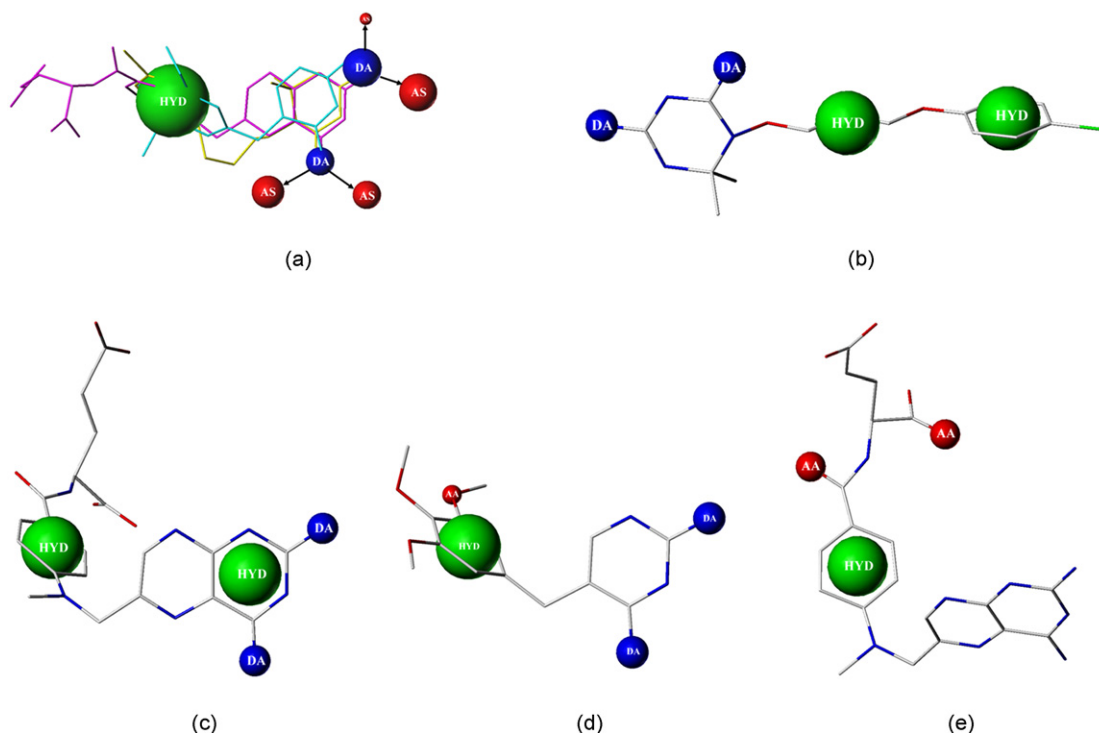


Fig. 2. Five pharmacophore hypotheses generated using three MtDHFR inhibitors.

**Table 1**  
Statistical parameters from screening *MtDHFR* test set molecules

Parameter	H1	H2	H3	H4	H5
Total number of molecules in the database ( <i>D</i> )	1285	1285	1285	1285	1285
Total number of actives in the database ( <i>A</i> )	78	78	78	78	78
Total hits ( <i>H<sub>t</sub></i> )	75	17	13	14	756
Active hits ( <i>H<sub>a</sub></i> )	74	17	12	11	37
% Yield of actives ( $H_a/H_t \times 100$ )	98.66	100	92.30	78.57	0.04
% Ratio of actives ( $H_a/A \times 100$ )	94.87	21.79	15.38	14.10	47.43
Enrichment factor ( <i>E</i> ) <sup>a</sup>	16.25	16.47	15.20	12.94	0.80
False negatives ( <i>A</i> – <i>H<sub>a</sub></i> )	4	61	66	67	41
False positives ( <i>H<sub>t</sub></i> – <i>H<sub>a</sub></i> )	1	0	1	3	719
Goodness of hit score ( <i>G</i> ) <sup>b</sup>	0.9763	0.8044	0.7301	0.6229	0.0627

<sup>a</sup>  $E = H_a \times D / H_t \times A$ .

<sup>b</sup>  $G = (H_a/4 \times H_t \times A) / (3A + H_t) \times (1 - [(H_t - H_a)/(D - A)])$ .

The validity of any pharmacophore model needs to be determined by applying that model to the test set to find out how correctly the model predicts the activity of the test set molecules and, most importantly, whether it can identify active and inactive molecules correctly. The pharmacophores were validated by accessing the predictive ability of the pharmacophore on test set database consisting of 78 known inhibitors of *MtDHFR* collected from literature and a subset of World of Molecular Bioactivity database consisting of 1207 molecules, active against different proteins other than used in present study, considered here as inactive. This validation gives confidence to select the best pharmacophore among the five pharmacophores generated. The results for pharmacophore validation are summarized in Table 1. A number of parameters such as hit list (*H<sub>t</sub>*), number of active percent of yields (%Y), percent ratio of actives in the hit list (%A), enrichment factor (*E*), false negatives, false positives and goodness of hit score (*G*) are calculated (Table 1) while carrying out the pharmacophore-based screening of test set molecules. The pharmacophore hypothesis derived using the GASP program (Hypothesis 1 or H1 in Table 1), with minimal false positives and negatives, good enrichment factor and goodness of fit score was considered as best model for virtual screening among the five pharmacophore hypotheses. In 75 molecules predicted to be active, 74 molecules were correctly picked thus missing only 4 false negatives with only one false positive. In pharmacophore Hypothesis 2 (H2) no false positive was found but only 17 actives were picked among the 78 actives that is why pharmacophore Hypothesis 1 was preferred over hypothesis number 2 (Table 1).

### 3.2. Virtual screening

We have also analyzed the physicochemical profile of the Maybridge small molecule database consisting of 59275 compounds used in the present virtual screening experiment. We focused on three properties/descriptors, molecular weight, number of rotatable bonds and log *P* values [31–33]. As seen in Fig. 3, Maybridge collection has values for these 3 properties consistent with drug-like rules. For 97.17% of the compounds, the MW is between 100 and 500 Da, with 31.83% of the compounds having a MW between 200 and 300 Da (Fig. 3a). Concerning the number of rotatable bonds, 94.16% of the compounds have between 0 (rigid) and 8 flexible bonds, with 57.35% of the compounds having less than five rotatable bonds (Fig. 3b). The calculated log *P* values range from –2 (more polar) to 5 (more hydrophobic) for 85.53% of the compounds (Fig. 3c). Thus, we believe that this Maybridge collection is sufficiently diverse in term of MW, log *P* and number of rotatable bond and suitable for a

virtual screening study. Overall, the active compounds display similar properties as the other molecules present in Maybridge database.

The best-validated Hypothesis H1, pharmacophore hypothesis derived using GASP program was used to screen the Maybridge small molecule database consisting of 59275 molecules. Virtual screening was carried out using flex utility of UNITY module available with SYBYL7.1. Pharmacophore-based virtual screening with yielded 632 hits that met the specified requirements.

### 3.3. Molecular docking and scoring

Molecular docking was then performed on 78 known inhibitors of *MtDHFR* and 632 virtual screening hits in the binding site of *MtDHFR*. The hydrophobic binding pocket of *MtDHFR* is made up of key residues Ile5, Trp6, Asp27, Gln28, Phe31, Arg32, Arg60, Ile94 and Tyr100. *MtDHFR*–NADPH–MTX ternary complex (PDBid-1DF7) 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. In case of known *MtDHFR* inhibitors only best scoring pose for each ligand was further selected. The best pose was determined using CScore module of SYBYL7.1 which generates the consensus of FlexX\_Score and four other scoring functions (G\_Score, PMF\_Score, D\_Score and ChemScore). The docked pose with highest CScore and FlexX\_Score value was finally selected as best pose for known *MtDHFR* inhibitors. For virtual screening hits retrieved using pharmacophore Hypothesis H1 only those hits were selected for which the top FlexX scoring pose had good FlexX score (–10 to –35 kJ/mol) 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 displayed good binding mode. Finally 183 virtual hits and 78 known inhibitors of *MtDHFR* in total were taken as promising candidates for further investigation.

### 3.4. SIFT-based analysis of docking result, similarity analysis and clustering to identify leads

The main objective of our study was to integrate the utility of structure interaction fingerprints as virtual screening tools when they are applied to the same target proteins. The structure interaction fingerprints were generated for 78 inhibitors of *MtDHFR* and 183 virtual screening hits selected after molecular docking and scoring using the method described above. The best poses selected after the molecular docking studies of the virtual screening hits and known inhibitors were clustered using the various implementations of the structure interaction

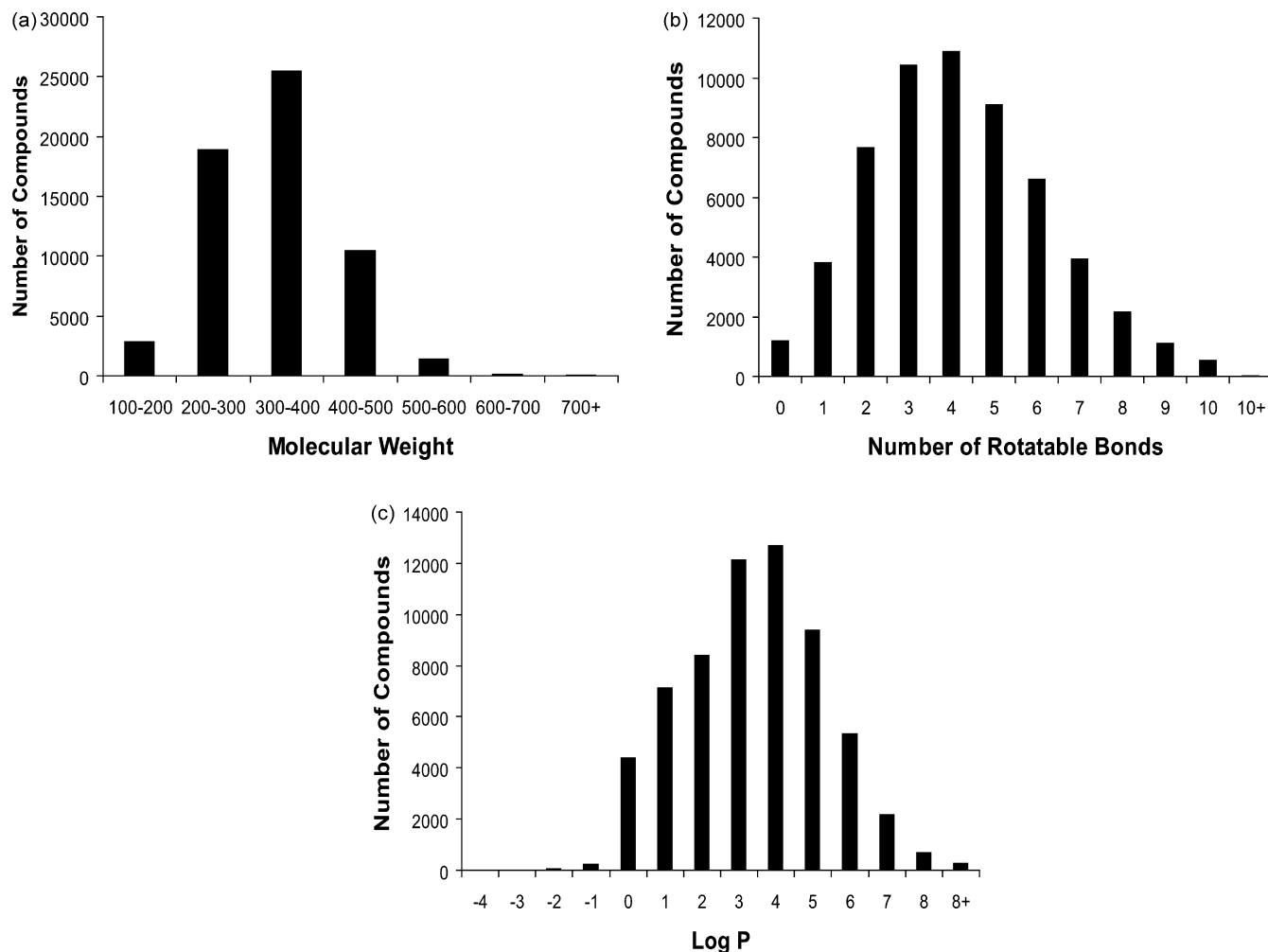


Fig. 3. Physicochemical profile of the Maybridge 59275-compound collection: (a) molecular weight; (b) number of rotatable bonds; (c) log *P*.

fingerprint method described above in Section 2. The approach involves computing the similarity metric (Rogers and Tanimoto coefficient) between all of the fingerprints and then using a hierarchical clustering algorithm to group similar interaction fingerprints together. The dendrogram derived by clustering

structure interaction fingerprints of *Mt*DHFR-inhibitor/virtual screening hits complex is shown in Fig. 4. The dendrogram clearly revealed two major clusters each of which represents a distinct binding pattern. Cluster 1 composed of 83.56% inhibitors of *Mt*DHFR (mostly highly active and medium active)

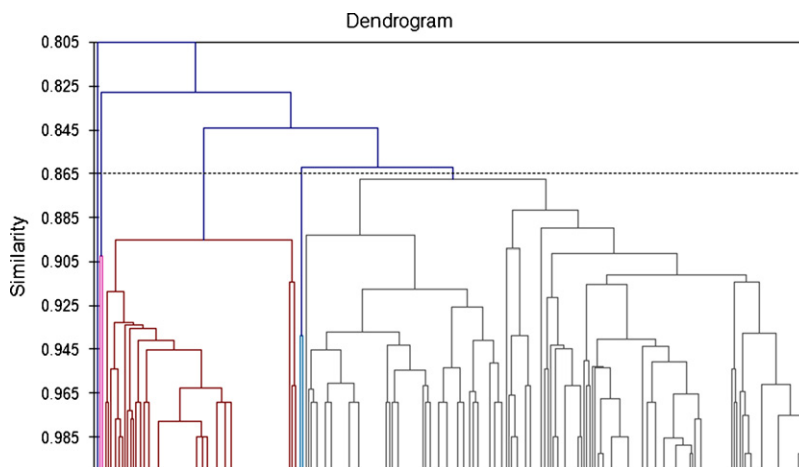


Fig. 4. Dendrogram derived by agglomerative hierarchical clustering of structure interaction fingerprints (SIFt) of *Mt*DHFR inhibitors and virtual screening hits.

and 65.58% virtual screening hits interacting with the protein. Similarly Cluster 2 is composed of 16.44% *MtDHFR* inhibitors mostly low active and 31.69% virtual screening hits. The remaining 2.73% of the virtual screening hits do not belong to these two major clusters and they are either singleton or form tiny clusters. To further evaluate the performance of SIFT-based clustering, the results were compared with chemical structure-based clustering. The maximum common substructure algorithm was applied to cluster the known inhibitors and virtual screening hits using LibMCS program [34]. Unlike the SIFT results where structurally diverse virtual screening hits are clustered together based on common receptor interactions, the chemical structure-based clustering groups structurally similar compounds (three inhibitors classes viz deazapteridine derivatives, triazine derivatives and pyrimidine 2,4-diamines) together and diverse compounds in different clusters.

It is noteworthy that the hierarchical clustering procedure, based solely on ligand receptor interaction features, is able to group structures into meaningful clusters where variable ligands have similar interactions with receptor and where very similar

ligands interact in a highly conserved way. It is interesting to note that, methotrexate, a well-known inhibitor of DHFR is clustered in Cluster 2 with inactive and the low active *MtDHFR* inhibitors that may be because of the interaction contributed by the glutamyl chain that is not present in Trimethoprim, WRB and other known inhibitors of *MtDHFR*.

We have used structure interaction fingerprints to examine conserved interactions between the ligands and *MtDHFR*. These conserved interactions can be used to understand critical determinant for inhibitor binding and selectivity and to prioritize potential hits from virtual screening. We have determined the degree of conservation of interactions *MtDHFR*–inhibitor complexes by exploring the frequency of contacts at each residue position. These interactions could be used as basic filter in virtual screening for the identification of potential hits of *MtDHFR*. Figs. 5 and 6 show that the clustering by their SIFT patterns has separated the best docked poses of *MtDHFR* inhibitors and virtual screening hits into different groups with distinct binding interactions. Interestingly, each of these clusters is comprised of poses having similar binding modes with the receptor; Cluster

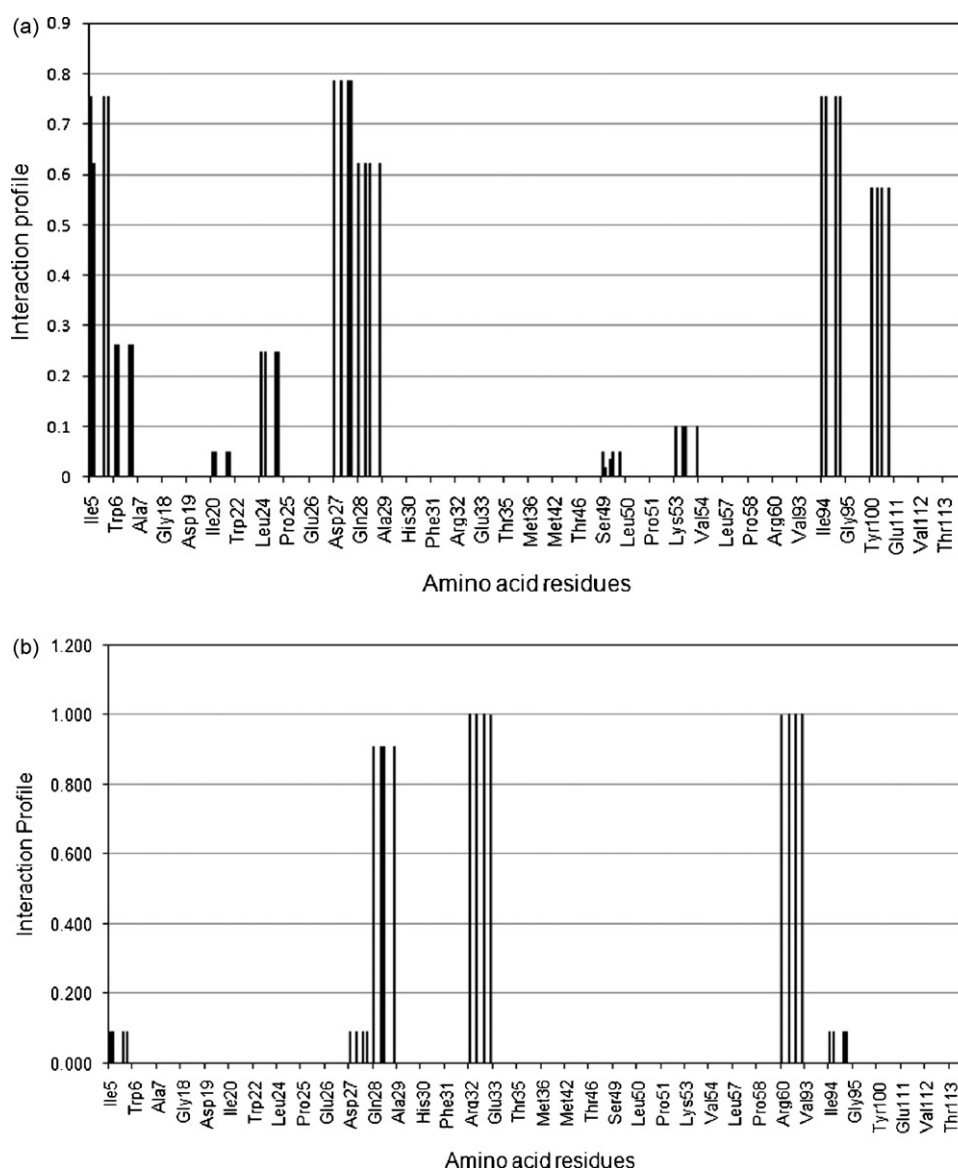


Fig. 5. Structure interaction fingerprint (SIFT) profile for *MtDHFR* inhibitors: (a) Cluster 1; (b) Cluster 2.



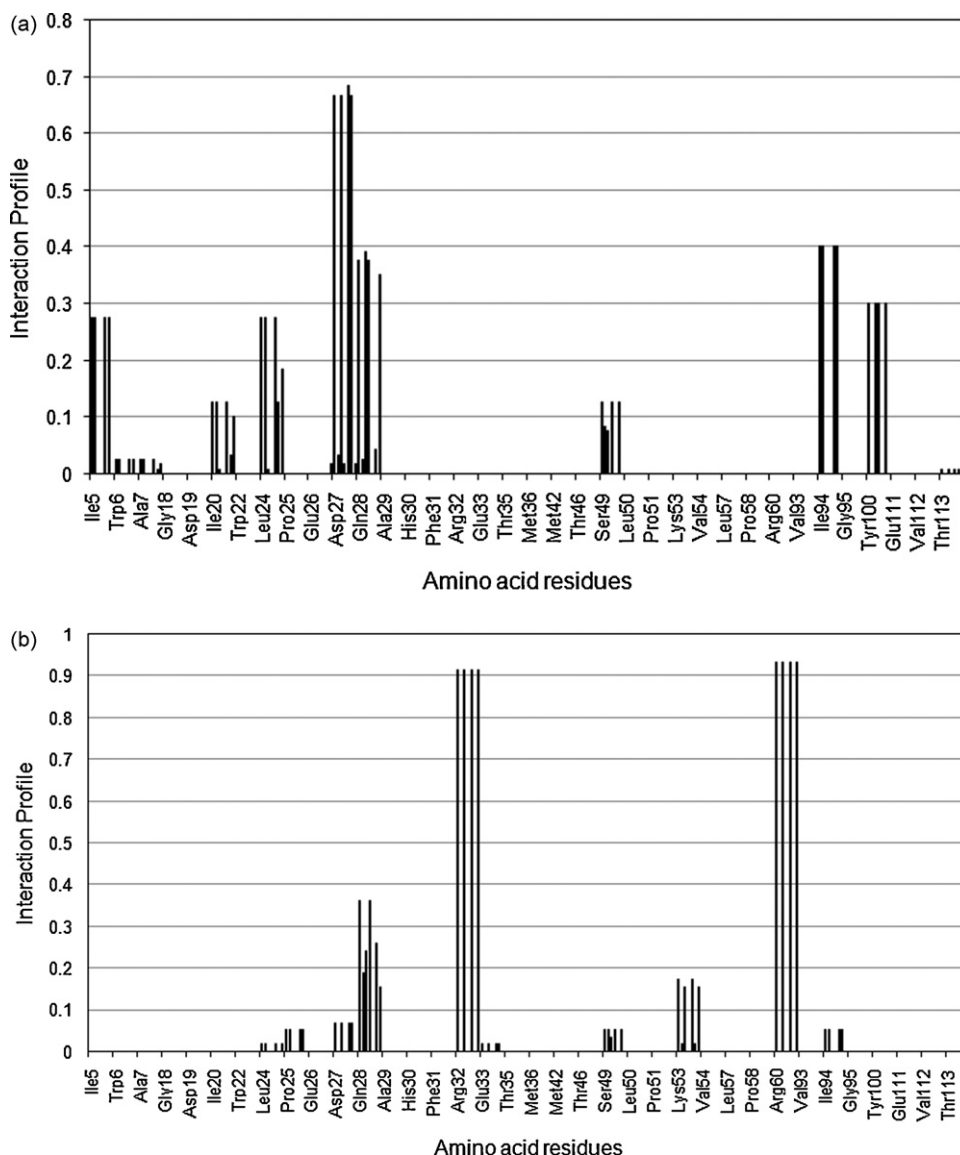


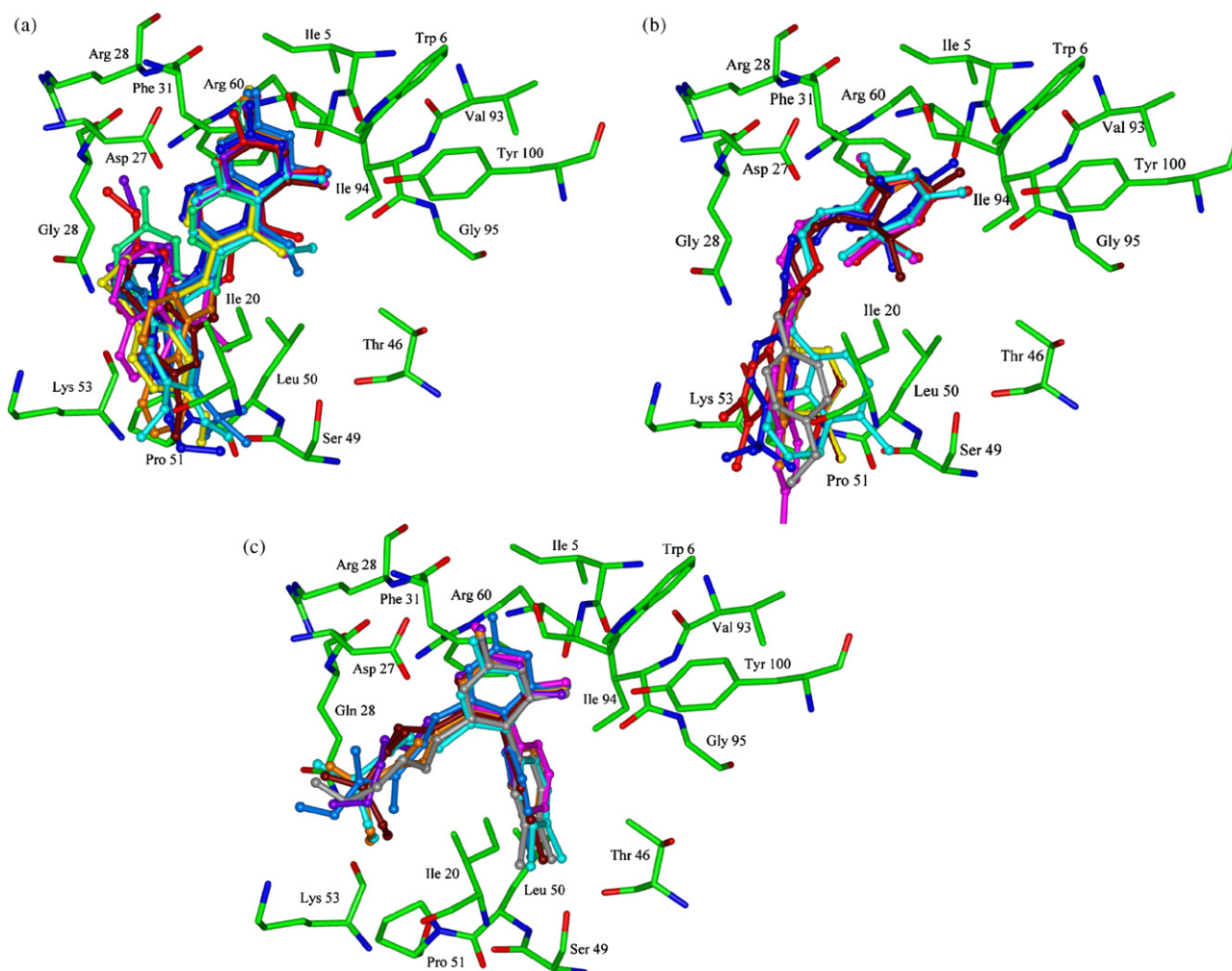
Fig. 6. Structure interaction fingerprint (SIFt) profile for virtual screening hits: (a) Cluster 1; (b) Cluster 2.

1 contains molecules more or less similar to the known X-ray crystal structures of *Mt*DHFR in complex with inhibitors. Cluster 2 compounds represent distinct binding mode that result in dissimilar interactions with the active site pocket formed by residues Ile5, Asp27, Ile94 and Tyr100. Finally, other tiny cluster compounds are outside the inhibitor binding site. The two major clusters represents two distinct binding modes as confirmed by careful inspections of their structures, although the results indicate that within each cluster there exists considerable variation in their interaction patterns. Agglomerative hierarchical clustering identifies distinct SIFt sub-clusters representing each cluster.

The virtual screening workflow followed, allows our method to be applied for the automated visualization of the putative binding modes using cluster analysis. After clustering the SIFts, one could trace the active clusters (cluster where the known actives or reference compound fall). This will enable the straight forward short listing of compounds and the binding modes as it would allow the users to focus attention on just the lead clusters. This approach can be especially valuable for compound selection in

large scale docking projects, which can field top-scoring sets containing tens of thousands compounds, but manual visualization of individual binding modes can still be a problem even with more manageable number of compounds. The results indicate that clustering can group all active compounds into just a few active clusters.

Compounds in Cluster 1 includes 83.56% of the known *Mt*DHFR inhibitors that bind to *M. tuberculosis*-DHFR via contacts with strands  $\beta$ -A,  $\beta$ -E and two helices,  $\alpha$ B and  $\alpha$ C. Three distinct classes of *Mt*DHFR inhibitors viz. deazapteridine derivatives, triazine derivatives and pyrimidine 2,4-diamines display overall similar yet unique binding signatures at the active site. Fig. 7 represents the binding mode of three different classes of *Mt*DHFR inhibitors viz. deazapteridine derivatives (Fig. 7a), triazine derivatives (Fig. 7b) and pyrimidines 2,4-diamines (Fig. 7c). Inhibitors bind essentially in the same pocket and their nitrogen containing heteroatomic ring position in the same orientation. A common feature of the binding mode is a set of strong hydrogen-bonding interactions between the heteroatomic rings and several structurally conserved residues. The heteroatomic ring of these



**Fig. 7.** Binding mode of MtDHFR inhibitors: (a) deazapteridine derivatives; (b) triazine derivatives; (c) pyrimidines 2,4-diamines.

inhibitors is situated into a mainly hydrophobic pocket and interacts strongly with the protein through four hydrogen bonds, side chain oxygen of Asp27 forms hydrogen bond with one amino group and sometimes with ring nitrogen and this particular interaction is highly conserved throughout DHFR of various species. Ile5 of strand  $\beta$ -A and Ile94 in the carboxy terminal part of  $\beta$ -E form a hydrogen bond each between the main chain carbonyl oxygen atom and other amino group of the inhibitors. In addition the side chain hydroxyl of Tyr100 in helix  $\alpha$ F also forms hydrogen bond with amino group of most of the active compounds. This arrangement of hydrogen bonds around the nitrogen containing ring is highly conserved among the known DHFR structures from various species. In addition to the hydrogen-bonding interactions, there are many van der Waals contacts between the inhibitors and the protein where the deazapteridine, triazine or pyrimidine ring is in contact with Ile5, Trp6, Gln28, Phe31 and Ile94. Most of the DHFR inhibitors interact with protein mainly via hydrogen-bonding and van der Waals interaction with the heteroatomic ring, however, the phenyl side chain of these inhibitors corresponding to hydrophobic feature of our pharmacophore hypothesis extends out towards the entrance of the binding pocket and interacts with the protein mainly by hydrophobic contacts with Ile20, Leu50 and Pro51.

In order to select promising compounds through visual inspection of the clusters and compound binding mode, following

criteria was considered: (1) compound should be from Cluster 1; (2) degree of occupancy of the enzyme in particular related to protein ligand surface complementarities; (3) formation of hydrogen bond with Asp27 and should form hydrogen bond with at least two of amino acid residues Ile5, Ile94 and Tyr 100; (4) hydrophobic contacts with Ile20, Leu50 and Pro51; (5) quality of the overall binding conformation from which 20 compounds have been selected as potential lead candidates.

The structures of the selected compounds are shown in Table 2, together with their corresponding molecular weight, calculated log *P* values. The promising compounds share common pharmacophoric features like a nitrogen-rich scaffold (triazine, pyrazolo-triazine, amides, piperazine, etc.) and hydrophobic group connected to the scaffold by linker groups of variable length. However, Tanimoto similarity coefficients estimated from MACCS KEYS indicate that the investigated compounds (1–20) are rather diverse.

The predicted binding modes of Maybridge hit KM 02351 and GK 03028 are described in Fig. 8. Like most of the MtDHFR inhibitors, identified hits are all anchored to the cavity by hydrogen bonds between the ligand and Ile5, Asp27, Ile94 and Tyr100. Aromatic rings of compounds KM 02351 (Fig. 8a) and GK 03028 (Fig. 8b) are embedded in the subpocket formed by Ile 20, Leu50 and Pro51 and interact with these residues via hydrophobic interactions.



**Table 2**Chemical structures of promising hits and their respective docking scores, molecular weight and log *P* identified using virtual screening

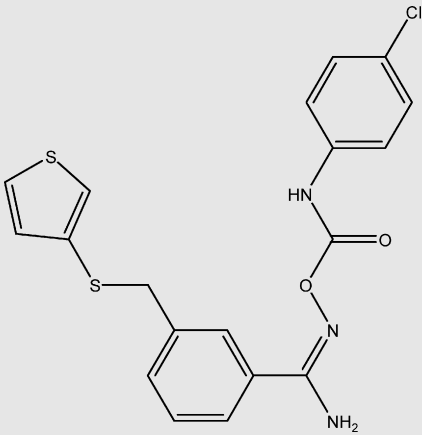
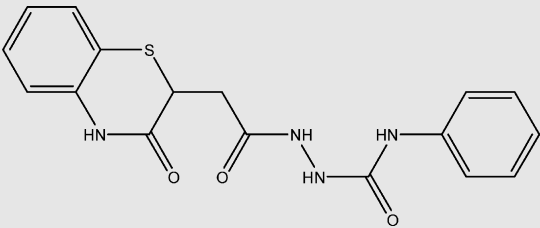
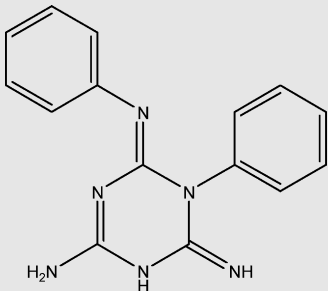
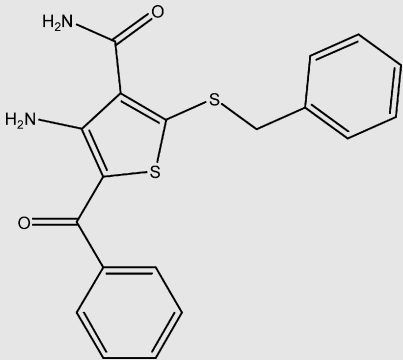
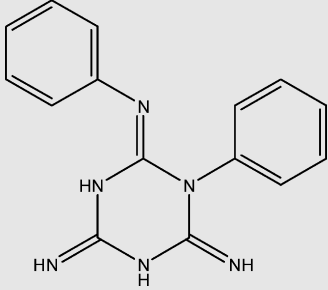
S. no.	Maybridge code	Structures	Flexx Dockscore (KJ/mol)	Molecular weight	log <i>P</i>
1	KM 02351		−24.443	417.94	6.055
2	KM 03725		−23.076	356.40	1.594
3	JFD 00933		−22.547	278.31	3.460
4	CD 01352		−22.392	368.48	3.808
5	NRB 03986		−20.901	278.31	3.883

Table 2 (Continued)

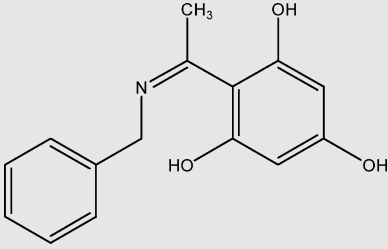
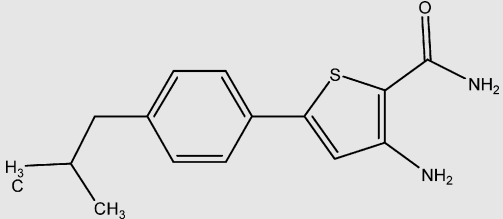
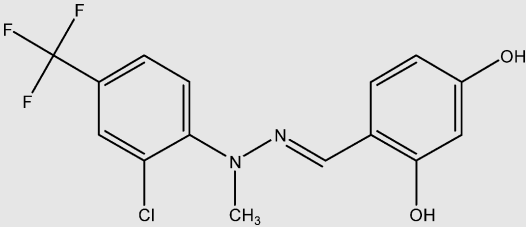
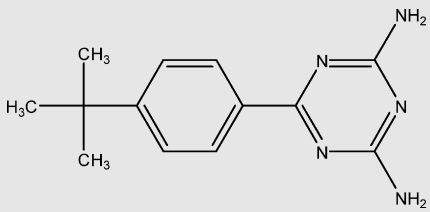
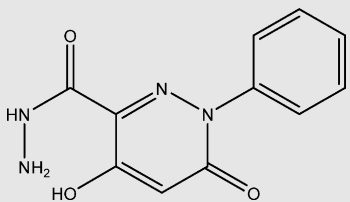
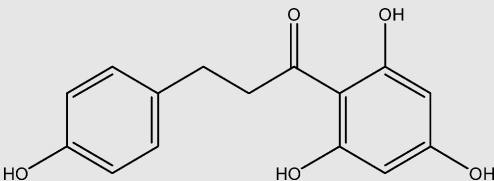
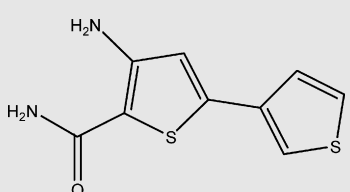
S. no.	Maybridge code	Structures	Flexx Dockscore (KJ/mol)	Molecular weight	log P
6	S 04627		–19.974	257.29	2.744
7	GK 03028		–19.077	274.38	3.154
8	SEW 02357		–19.041	345.71	4.607
9	RDR 02761		–18.930	243.31	1.357
10	BTB 12485		–18.647	246.22	0.188
11	RJC 02792		–18.141	274.27	2.435
12	GK 01140		–18.135	224.30	1.231

Table 2 (Continued)

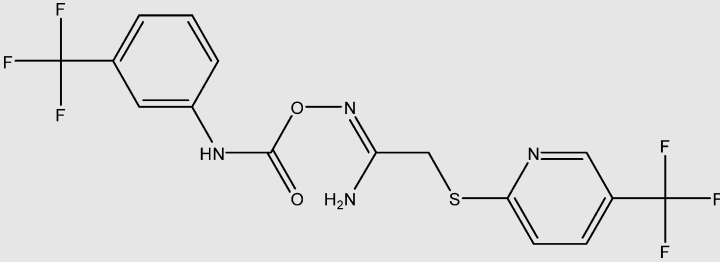
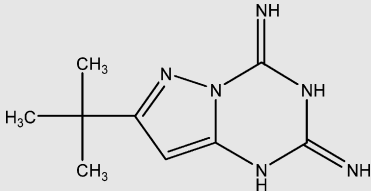
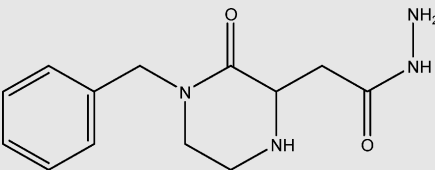
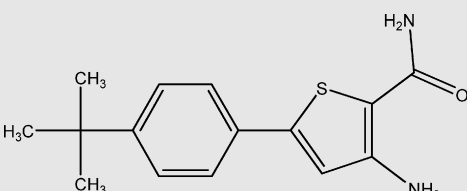
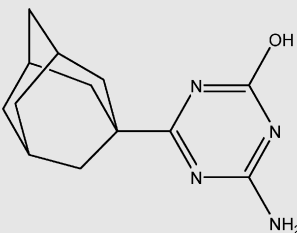
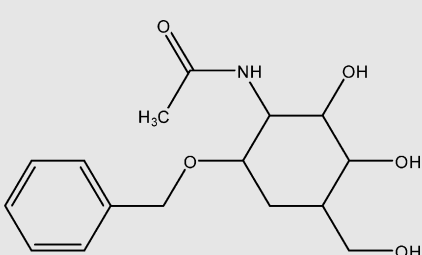
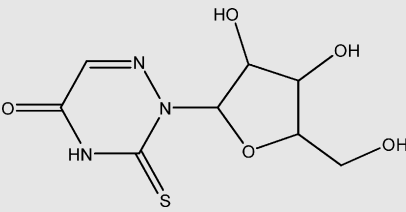
S. no.	Maybridge code	Structures	Flexx Dockscore (KJ/mol)	Molecular weight	log P
13	SPB 03064		−17.677	438.35	4.681
14	BTB 09381		−17.592	206.25	1.689
15	DP 01702		−17.530	262.31	−0.775
16	GK 03408		−17.252	274.38	3.080
17	RF 02819		−16.177	246.31	0.679
18	BTB 11574		−15.113	311.33	−0.585
19	BTB 13930		−14.933	261.26	−1.966

Table 2 (Continued)

S. no.	Maybridge code	Structures	Flexx Dockscore (KJ/mol)	Molecular weight	log P
20	BTB 11967		−10.480	292.37	0.845

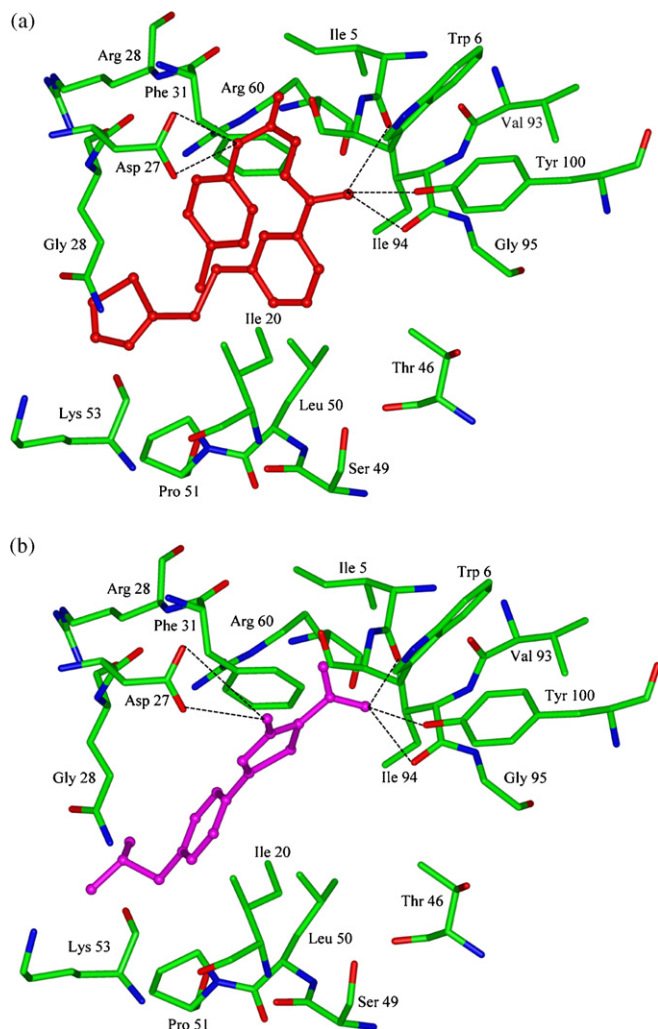


Fig. 8. Binding mode of potential hits: (a) KM 02351; (b) GK 03028.

#### 4. Conclusion

Here, we have applied a multi-step virtual screening protocol including 3D pharmacophore search, molecular docking and hierarchical clustering of structure interaction fingerprints to prioritize the virtual screening hits against *MtDHFR*, a potential therapeutic target against tuberculosis. Use of structural interaction fingerprints makes it possible to quickly shortlist promising compounds and automated visualization of putative binding modes using hierarchical clustering. It is encouraging to see that fingerprint scoring, though favoring only compounds that form

predetermined binding modes, is still able to suggest highly diverse sets of compounds, as indicated by the low Tanimoto self-similarity values obtained for the top-ranking hits studied here. The criterion for selection of promising hits was based on predicted binding affinities, i.e. scores, as well as binding interactions of the ligands as determined by FlexX. Finally, 20 molecules were considered to be the promising hits for further investigation against *MtDHFR*. This study should provide further insights to support structure-based design of anti-tubercular drugs to develop novel and diverse dihydrofolate reductase inhibitors with improved activity profiles.

#### Acknowledgements

This manuscript is CDRI communication number 7483. This work was supported by the Council of Scientific and Industrial Research (India) funded network project NWP0034 (Validation of identified screening models and development of new alternative models for evaluation of new drug entities). Ashutosh Kumar thanks CSIR for fellowship.

#### References

- [1] I.M. Kompis, K. Islam, R.L. Then, *Chem. Rev.* 105 (2005) 593–620.
- [2] J.M. Blaney, C. Hansch, C. Silipo, A. Vittoria, *Chem. Rev.* 84 (1984) 333–407.
- [3] S.J.B. Tendler, M.D. Threadgill, M.J. Tisdale, *Cancer Lett.* 36 (1987) 65–69.
- [4] K.A. Zanetti, P.J. Stover, *J. Biol. Chem.* 278 (2003) 10142–10149.
- [5] R.L. Then, *J. Chemother.* 16 (2004) 3–12.
- [6] W.J. Suling, R.C. Reynolds, E.W. Barrow, L.N. Wilson, J.R. Piper, W.W. Barrow, *J. Antimicrob. Chemother.* 42 (1998) 811–815.
- [7] R. Li, R. Sirawaraporn, P. Chitnumsub, W. Sirawaraporn, J. Wooden, F. Athappilly, S. Turley, W.G. Hol, *J. Mol. Biol.* 295 (2000) 307–323.
- [8] W.J. Suling, J.A. Maddry, *J. Antimicrob. Chemother.* 47 (2001) 451–454.
- [9] A.B. Gerum, J.E. Ulmer, D.P. Jacobus, N.P. Jensen, D.R. Sherman, C.H. Sibley, *Antimicrob. Agents Chemother.* 46 (2002) 3362–3369.
- [10] M.H. El-Hamamsy, A.W. Smith, A.S. Thompson, M.D. Threadgill, *Bioorg. Med. Chem.* 15 (2007) 4552–4576.
- [11] J.A. DiMasi, R.W. Hansen, H.G. Grabowski, *J. Health Econ.* 22 (2003) 151–185.
- [12] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Res.* 28 (2000) 235–242.
- [13] P. Ertl, *J. Chem. Inf. Comput. Sci.* 43 (2003) 374–380.
- [14] W.P. Walters, M.T. Stahl, M.A. Murcko, *Drug Discov. Today* 3 (1998) 160–178.
- [15] G.L. Warren, C. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, *J. Med. Chem.* 49 (2006) 5912–5931.
- [16] Z. Deng, C. Chuaqui, J. Singh, *J. Med. Chem.* 47 (2004) 337–344.
- [17] C. Chuaqui, Z. Deng, J. Singh, *J. Med. Chem.* 48 (2005) 121–133.
- [18] Z. Deng, C. Chuaqui, J. Singh, *J. Med. Chem.* 49 (2006) 490–500.
- [19] J. Singh, Z. Deng, G. Narale, C. Chuaqui, *Chem. Biol. Drug. Des.* 67 (2006) 5–12.
- [20] Sybyl, Version 7.1, Tripos, Inc., St. Louis, MO, 2005.
- [21] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* 261 (1996) 470–489.
- [22] G. Jones, P. Willett, G. Glen, *J. Mol. Biol.* 245 (1995) 43–53.
- [23] I. Muegge, Y.C. Martin, *J. Med. Chem.* 42 (1999) 791–804.
- [24] E.C. Meng, B.K. Shoichet, I.D. Kuntz, *J. Comput. Chem.* 13 (1992) 505–524.
- [25] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, *J. Comput. Aided Mol. Des.* 11 (1997) 425–445.
- [26] I.K. McDonald, J.M. Thornton, *J. Mol. Biol.* 238 (1994) 777–793.
- [27] D.J. Rogers, T.T. Tanimoto, *Science* 132 (1960) 1115–1118.
- [28] R. Dubes, A.K. Jain, *Adv. Comput.* 19 (1980) 113–228.

- [29] SYSTAT, SYSTAT for Windows Version 12, SYSTAT Software Inc., Richmond, California, 2007.
- [30] INSIGHT II 2000.1, Molecular Modeling Software, Accelrys, Inc., San Diego, USA, 2000.
- [31] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv. Drug Deliv. Rev.* 46 (2001) 3–26.
- [32] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, *J. Med. Chem.* 45 (2002) 2615–2623.
- [33] R. Wang, Y. Gao, L. Lai, *Perspect. Drug Discov. Des.* 19 (2000) 47–66.
- [34] LibMCS Program Version 0.7, ChemAxon Ltd., Maramaros köz3/a 1037 Budapest, Hungary ([www.chemaxon.com](http://www.chemaxon.com)).