A Virtual Screening Approach for Thymidine Monophosphate Kinase Inhibitors as Antitubercular Agents Based on Docking and Pharmacophore Models

B. Gopalakrishnan,*,† V. Aparna,‡ J. Jeevan,‡ M. Ravi,‡ and G. R. Desiraju*,‡

Bioinformatics Division, Advanced Technology Centre, TATA Consultancy Services Limited, 1 Software Units Layout, Madhapur, Hyderabad 500 081, India, and School of Chemistry, University of Hyderabad, Hyderabad 500 046, India

Received February 22, 2005

Docking and pharmacophore screening tools were used to examine the binding of ligands in the active site of thymidine monophosphate kinase of *Mycobacterium tuberculosis*. Docking analysis of deoxythymidine monophosphate (dTMP) analogues suggests the role of hydrogen bonding and other weak interactions in enzyme selectivity. Water-mediated hydrogen-bond networks and a halogen-bond interaction seem to stabilize the molecular recognition. A pharmacophore model was developed using 20 dTMP analogues. The pharmacophoric features were complementary to the active site residues involved in the ligand recognition. On the basis of these studies, a composite screening model that combines the features from both the docking analysis and the pharmacophore model was developed. The composite model was validated by screening a database spiked with 47 known inhibitors. The model picked up 42 of these, giving an enrichment factor of 17. The validated model was used to successfully screen an in-house database of about 500 000 compounds. Subsequent screening with other filters gave 186 hit molecules.

INTRODUCTION

Computational procedures have become an integral component of most drug discovery programs. Structure-based virtual screening (VS) methods are emerging as reliable and complementary approaches to high throughput screening in the lead discovery process, from hit identification to lead optimization.[1] A variety of computational techniques that allow the reduction of enormous virtual libraries ($\sim 10^{12}$) to more manageable proportions constitute the VS protocols.[2] In the absence of any knowledge of the target, VS includes ligand-based screens such as 1D filters (e.g., molecular weight), 2D filters (similarity, substructure fingerprints[3]), and 3D filters (3D-pharmacophore,[4,5] 3D shape matching[6]). Some of these filters such as Lipinski's rule of five can be used as coarse virtual screens for drug likeness.[7] Tools based on 2D molecular descriptors cannot adequately gauge the molecular similarity of pairs of compounds that are structurally different but behave alike in binding to a target. However, if conformational flexibility is not considered, the pharmacophore method may miss some of the hits. Structure-based VS tools include receptor−ligand docking[8−10] and molecular dynamics. Docking, although computationally intensive, is emerging as a reliable high-resolution tool to study recognition between the ligand and its cognate receptor.[11−13]

Successful VS depends on the choice of screening tools and their implementation in each phase of the screen. The scope of the present work lies in the determination of the best VS tool and its application for the system of choice. The idea was to develop a robust VS model that can combine both the accuracy of docking and the speed of the pharmacophore method for the system of interest, namely, *Mycobacterium tuberculosis* thymidine monophosphate kinase ($TMPK_{mt}$).

Tuberculosis (TB), second only to AIDS among infectious diseases, kills more than 2 million people a year worldwide.[14] The emergence of multiple-drug-resistant TB and its synergism with HIV stimulated the search for new targets and the development of new drugs to overcome the global threat.[15] $TMPK_{mt}$ is one of the promising targets in the treatment of tuberculosis.[16] TMPK is an enzyme that catalyses the conversion of deoxythymidine monophosphate (dTMP) to deoxythymidine diphosphate (dTDP) using ATP as a phosphoryl donor.[17] TMPK inhibitors have gained much attention because of their prospects as antimetabolites. Also, these molecules can be promising leads to treat tuberculosis by inhibiting DNA synthesis in *M. tuberculosis*. Very few lead generation and optimization studies have been reported for $TMPK_{mt}$ inhibitors;[18−20] the availability of a high-resolution crystal structure[21] is handy for comparing different approaches in the present study and subsequently developing a composite and robust VS model.

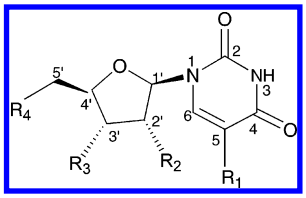
MATERIALS AND METHODS

Selection of Molecules. A set of 47 compounds reported as thymidine monophosphate kinase inhibitors was compiled (Table 1).[19,20,22,23] The inhibitory activities were converted into the corresponding pK_i ($-\log K_i$) values, where the K_i value represents the drug concentration that causes inhibition of dTMP phosphorylation by $TMPK_{mt}$. All of the K_i values had been obtained using the same assay method.[19] pK_i values of all of the molecules spanned a sufficiently wide range from 2 to 6.

* Corresponding authors. Tel.: +91 40 23134828; fax: +91 40 23010567; e-mail: gautam_desiraju@yahoo.com (G.R.D.). Tel.: +91 40 55673591; fax: +91 40 55672222; e-mail: gopal@atc.tcs.co.in (B.G.).
† TATA Consultancy Services Limited.
‡ University of Hyderabad.

10.1021/ci050064z CCC: $30.25
Published on Web 05/27/2005

Table 1. Actual and Predicted Activities of Training and Test Set Molecules Obtained from Pharmacophore



molecule number	R_1	R_2	R_3	R_4	actual activity pK_i	predicted activity pK_i
1	CH_3	H	OH	$NHCOCH_3$	4.045	5.200[b]
2	CH_3	H	OH	N_3	5.154	5.096[a]
3	CH_3	H	OH	NH_2	4.920	5.045[a]
4	CH_3	OH	OH	OH	3.145	4.011[b]
5	CH_3	OH[c]	OH	OH	3.623	4.107[b]
6	CF_3	H	OH	OH	4.013	3.698[a]
7	C_2H_5	H	OH	OH	2.940	3.309[a]
8	F	OH	OH	OH	3.283	3.318[a]
9	F	OH	OH	H	3.251	3.229[b]
10	CH_3	H	NH_2	OH	3.638	3.468[b]
11	CH_3	H	F	OH	4.552	3.327[b]
12	CH_3	F	OH	OH	3.913	4.387[b]
13	CH_3	OH	N_3	I	3.309	4.823[b]
14	CH_3	OH	NH_2	H	2.721	2.920[a]
15	CH_3	OH	$NHC(NH)NH_2$	OH	3.853	3.301[b]
16	CH_3	H	CH_2N_3	OPO_3^{2-}	4.920	3.327[b]
17	CH_3	H	CH_2NH_2	OPO_3^{2-}	4.978	3.638[b]
18	CH_3	H	CH_2F	OPO_3^{2-}	4.823	3.522[b]
19	CH_3	H	CH_2OH	OPO_3^{2-}	4.537	3.568[b]
20	CH_3	OH	CH_2N_3	OPO_3^{2-}	3.933	3.795[a]
21	CH_3	OH	CH_2NH_2	OPO_3^{2-}	3.501	3.677[b]
22	CH_3	H	CH_2N_3	OH	4.397	3.823[b]
23	CH_3	H	CH_2NH_2	OH	4.244	3.318[b]
24	CH_3	H	CH_2F	OH	4.346	3.301[b]
25	CH_3	H	CH_2OH	OH	4.387	4.522[a]
26	CH_3	OH	CH_2N_3	OH	3.113	3.318[a]
27	CH_3	OH	CH_2NH_2	OH	3.405	3.327[a]
28	CH_3	H	CH_2CH_2OH	OH	3.806	4.000[a]
29	CH_3	H	NH_2	OPO_3^{2-}	3.628	3.552[b]
30	CH_3	OH	NH_2	OPO_3^{2-}	4.568	3.455[b]
31	CH_3	NH_2	OH	OPO_3^{2-}	4.259	4.537[a]
32	CH_3	Cl	OH	OPO_3^{2-}	4.721	4.119[b]
33	CH_3	F	OH	OPO_3^{2-}	4.366	4.494[a]
34	$C_6H_5CH_2$	H	OH	OPO_3^{2-}	4.552	3.638[b]
35	CH_3	H	N_3	OPO_3^{2-}	5.000	4.853[a]
36	CH_3	H	N_3	OH	4.552	4.657[b]
37	Br	H	N_3	OH	4.978	4.886[b]
38	Br	H	OH	OH	5.301	5.221[a]
39	CH═CHBr	H	OH	OH	3.204	4.468[b]
40	CH_2OH	H	OH	OH	3.086	3.244[b]
41	Cl	H	N_3	OH	4.795	4.657[a]
42	CH_3	H	OH	OH	4.568	4.408[a]
43	H	H	OH	OH	2.991	3.070[a]
44	F	H	OH	OH	3.673	3.301[a]
45	I	H	OH	OH	4.481	4.568[a]
46	OH	H	OH	OH	3.568	3.346[a]
47	H	H	N_3	OH	3.091	3.251[b]

[a] Training set. [b] Test set. [c] Hydroxyl in β position.

Molecular Modeling. All of the molecules were built from the coordinates of the dTMP structure (1G3U) using the builder module of Cerius2.[24] All of the structures were minimized using the steepest descent algorithm with a convergence gradient value of 0.001 kcal/mol. Partial atomic charges were calculated using the Gasteiger method. Further geometry optimization was carried out for each compound with the MOPAC 6 package using the semiempirical AM1 Hamiltonian.

Docking. The crystal structure of $TMPK_{mt}$ (1G3U) with its bound substrate dTMP was used in the study. Hydrogen atoms were added to the protein while keeping all of the residues in their charged form. Initially, all of the hydrogens were minimized, keeping all heavy atoms fixed. The whole complex including the bound water was energy minimized by the steepest descent followed by conjugate gradient methods to attain a convergence gradient of 0.01 kcal/mol using CVFF in InsightII.[25] All of the amino acids within a 13 Å radius from the center of the bound ligand were considered to constitute the active site.

Docking was carried out using GOLD,[26] which uses the genetic algorithm (GA).[27] For each of the 10 independent

GA runs, a maximum number of 100 000 GA operations were performed on a set of five groups with a population size of 100 individuals. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. Default cutoff values of 2.5 Å (d_{H-X}) for hydrogen bonds and 4.0 Å for van der Waals were employed. Hydrophobic fitting points were calculated to facilitate the correct starting orientation of the ligand for docking by placing the hydrophobic ligand atoms appropriately in the corresponding areas of the active site. When the top three solutions attained root-mean-square deviation (rmsd) values within 1.5 Å, GA docking was terminated. The first ranked solutions of the ligands were taken for further analysis.

Pharmacophore Modeling. Multiple acceptable conformations were generated for all ligands within the Catalyst[28] ConFirm module using the "Poling" algorithm.[29] A maximum of 250 conformations were generated for each molecule within an energy threshold of 20.0 kcal/mol above the global energy minimum. The training set molecules (20) associated with their conformations were submitted to the Catalyst hypothesis generation (HypoGen) (Table 1). Features such as hydrogen-bond donor (HBD), hydrophobic (HY), and negative ionizable features were included for the pharmacophore generation on the basis of common features present in the study molecules. The statistical parameters such as cost values determine the significance of the model (for more details on cost values, see ref 30). Ten pharmacophore models with significant statistical parameters were generated. The best model was selected on the basis of a high correlation coefficient (r), lowest total cost, and rmsd values. The final model was further validated by a test set of 27 molecules.

VIRTUAL SCREENING

Database Generation. Various publicly available databases, namely, IBS (90 000),[31] ChemStar (20 000),[32] MDPI (10 655),[33] bionet (43 176),[34] and Asinex (39 000),[35] were compiled. For each molecule in the database, 100 conformers were generated using the "fast fit" method in Catalyst. Other databases in the Catalyst module, including NCI 2000 (238 819),[36] Maybridge (55 273),[37] and MiniMaybridge (2000),[37] were also included to build a combined database of about 500 000 molecules.

Database Screening. The pharmacophore model developed using HypoGen was used as a 3D structural query in the VS. For this purpose, an additional HBD feature was added using the "merge hypothesis" option in Catalyst. This feature appeared in models 4 and 6 of the top 10 hypotheses (Table 2) corresponding to the N_1 of the thymidine ring in the ligand, which forms a hydrogen bond with the side-chain carbonyl of Asn100. This composite pharmacophore model (screen 1) was used for database searching by the "best flexible search" method in Catalyst. The hits obtained were further filtered using Lipinski's rule of five (screen 2) and the fitness score $\geq$ 3.00 (screen 3). The fingerprints of the hit molecules were used for cluster analysis using the Tanimoto similarity index. The hits were clustered into any particular group if the similarity between the fingerprints was $\geq$ 60%.

RESULTS AND DISCUSSION

The main objective of the present work is to develop a model for the VS of $TMPK_{mt}$ inhibitors. To accomplish this, we have used pharmacophore modeling and docking. A thorough analysis of each of the two methods was carried out to develop a composite screening model. The model is further used for screening an in-house database to identify new leads for $TMPK_{mt}$ inhibitors.

Table 2. Statistical Details of the Pharmacophore Models

hypothesis[a]	total cost	rmsd	correlation coefficient	features[b]
1	78.68	0.66	0.963	HBA, HBA, HBA, HY
2	80.35	0.75	0.951	HBA, HBA, HBA, HY
3	86.74	1.11	0.893	HBA, HBA, HBA, HY
4	87.05	1.12	0.890	HBA, HBA, HBD, HY
5	87.56	1.14	0.885	HBA, HBA, HBA, HY
6	87.87	1.16	0.882	HBA, HBA, HBD, HY
7	87.00	1.16	0.882	HBA, HBA, HBA, HY
8	88.59	1.16	0.881	HBA, HBA, HBA, HY
9	89.15	1.21	0.870	HBA, HBA, HBA, HY
10	89.63	1.23	0.865	HBA, HBA, HBA, HY

[a] Fixed cost = 74.302, configuration cost = 15.117, null cost = 119.075. All cost values are in bits. [b] Hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic feature (HY).

Docking. The substrate, dTMP, was docked into the active site of $TMPK_{mt}$. The rmsd between the docked pose of dTMP and its bound conformation in the crystal structure (1G3U) is 0.2 Å, indicating that GOLD was able to reproduce the correct pose. The detailed interactions of dTMP−$TMPK_{mt}$ are shown in Figure 1.

It can be seen that binding forces between dTMP and the enzyme manifest in (i) a stacking interaction between the thymidine ring and Phe70 (3.70 Å), (ii) a multicenter (bifurcated) hydrogen bond between the carbonyl O atom at thymidine C_4 and the terminal guanidino nitrogen atoms of Arg74 ($N_\omega H\cdots O{=}C$, 1.99 Å; $N_{\omega'}H\cdots O{=}C$, 1.92 Å; all distances are for d_{H-X}), (iii) a hydrogen bond between the $C_{3'}$−OH of dTMP and the side-chain carbonyl of Asp9 ($C_\delta{=}O\cdots HO$, 1.97 Å), (iv) a hydrogen bond between the phosphate oxygen of OP_2 and the guanidino N atom of Arg95 ($N_\delta H\cdots O{=}P$, 2.09 Å), (v) an ionic interaction between OP_1 and Mg^{2+} ($Mg^{2+}\cdots{}^{-}O{-}P$, 2.33 Å) where OP_1 occupies the coordination site of Mg^{2+}, (v) a hydrogen bond between the side-chain carbonyl oxygen of Asn100 and thymidine N_3 ($C_\delta{=}O\cdots HN$, 1.99 Å), and (vi) a weak hydrogen bond between the C_β of Tyr103 and the carbonyl oxygen at thymidine C_2 ($C_\beta H\cdots O{=}C_2$, 2.65 Å). The electrostatic interaction between OP_1 and Mg^{2+} is probably responsible for positioning the phosphate oxygen of dTMP in the active site for the kinase activity. All of the interactions observed above are in complete agreement with the crystal structure.[21] In addition to these interactions, the substrate dTMP is stabilized in the active site by a water-mediated finite σ-bond cooperative chain, OH (Tyr165)···OH (Wat1002)···OH (Wat1003)···OH (Ser104)···O=C (Asn100).

All of the molecules in the study were docked using the same method, and their interactions were analyzed. Only the interactions of the most- and least-active molecules are discussed in detail. The orientation and hydrogen bonds of the most-active molecule, **38**, within the active site of $TMPK_{mt}$ are shown in Figure 2.

Molecule **38** interacts with the active site residues similar to that of dTMP. Additionally, **38** also forms weak interactions like (i) a CH···N hydrogen bond with the side chain of Arg95 ($N_\delta\cdots HC_{5'}$, 2.43 Å; $N_\omega\cdots HC_{3'}$, 2.58 Å), (ii)

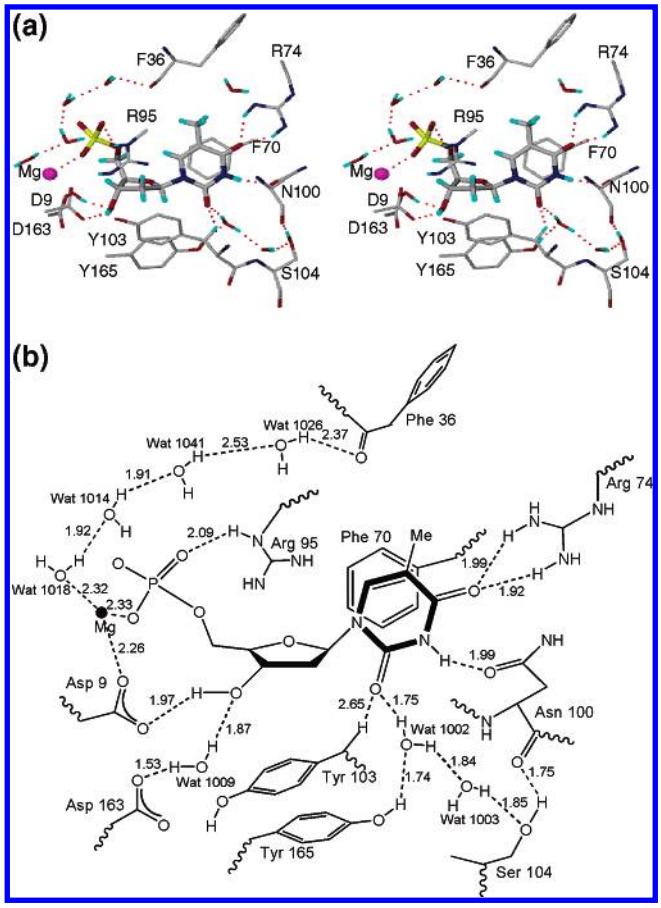


Figure 1. (a) Stereoview of the substrate dTMP−$TMPK_{mt}$ complex. Hydrogen bonds are shown as red dotted lines. Only those protein H atoms involved in hydrogen bonds are shown. Single letter amino acid codes are used for clarity. (b) Schematic representation of interactions of the substrate dTMP−$TMPK_{mt}$ complex in the active site. Hydrogen bonds are shown as dashed lines and are in Å.

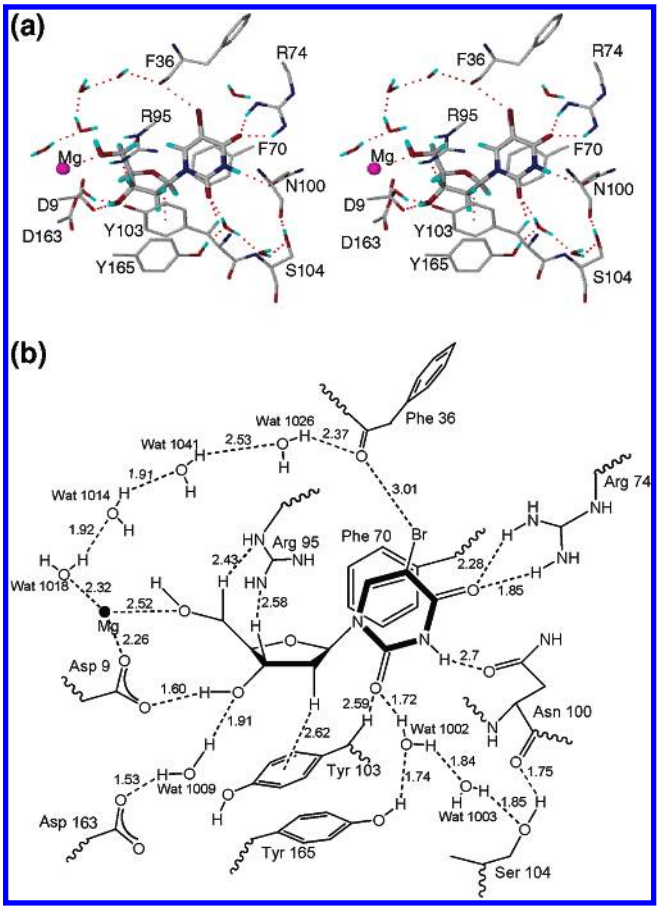


Figure 2. (a) Stereoview of the most-active molecule **38**−$TMPK_{mt}$ complex. (b) Schematic representation of the most-active molecule **38**−$TMPK_{mt}$ complex and the interactions within.

a CH···O hydrogen bond between the carbonyl oxygen at C_2 and the $C_\beta H$ of Tyr103 (2.59 Å), and (iii) a CH···π interaction between the π system of Tyr103 and the $C_{2'}$ of the ribose ring (2.62 Å). It has been reported that Tyr103 is crucial for the enzyme selectivity toward deoxyribonucleotides.[21] The weak interactions with Tyr103 appear to have a significant role in this selectivity.

Electrostatic interactions are observed between the hydroxyl oxygen atom at ribose $O_{5'}$ and Mg^{2+} (2.52 Å). The Br at thymine C_5 also forms a polarization-type contact with the backbone carbonyl oxygen of Phe36 ($Br^{\delta+}\cdots{}^{\delta-}O{=}C$, 3.01 Å). Replacing Br with a methyl group removes this interaction, resulting in a decrease in activity (Table 1). The halogen substitution at C_5 was mainly interpreted as a space filler.[16] However, the role of halogen at C_5 can be seen as the most stabilizing factor in the molecular recognition. The halogen here is not just a space filler but is involved in halogen bonding. Although this particular interaction is weak, its stabilization with the concomitant water-mediated cooperative hydrogen-bond network makes it worthwhile to increase the inhibitory activity prominently. Further exploration of such Br···O interactions using the Cambridge Structural Database (CSD; version 5.25, July 2004) revealed 580 hits within the distance range 2.8 Å < d (Br···O) < 4.0 Å, suggesting that such polarization-induced contacts are not uncommon in the crystal structures of small molecules.[38] Such interactions are stabilized to an extent of around 2 kcal/mol,[39] making them, in energetic terms, equivalent to C−H···O or C−H···N hydrogen bonds. Halogen···O interactions have not been reported often in analyses of macromolecular crystal structures.[40] A notable example, however, is the interaction of the iodine atoms of thyroid hormones with donor sites in the target protein.[41] In summary, the reader will note that in addition to the conventional or strong hydrogen bonds and electrostatic Mg^{2+}···O interactions, there are also several weaker interactions that, acting in concert, make a contribution to the binding specificity of the ligand in the active site. The orientation of the least-active molecule, **14**, is similar to that of the most-active molecule, **38**. However, fewer hydrogen bonds are present when compared to **38**, as shown in Figure 3. Also, the absence of Mg^{2+}···O electrostatic interactions and Br···O polarization-induced interactions are probably the reasons for the low activity.

It is worthwhile to mention that the ligand−protein complex is well-stabilized by two water-mediated finite σ-bond cooperative chains:[42] (i) OH (Wat1018)···OH (Wat1014)···OH (Wat1041)···OH (Wat1026)···O=C (Phe36) and (ii) OH (Tyr165)···OH (Wat1002)···OH (Wat1003)···OH (Ser104)···O=C (Asn100). The first network is extended to the ligand through the Br atom in **38** (Figure 2). The second network is common to all ligands. Such cooperative networks gain in energy, typically 20% over the isolated interactions. The presence of such water-mediated interactions of ligands within the enzyme active

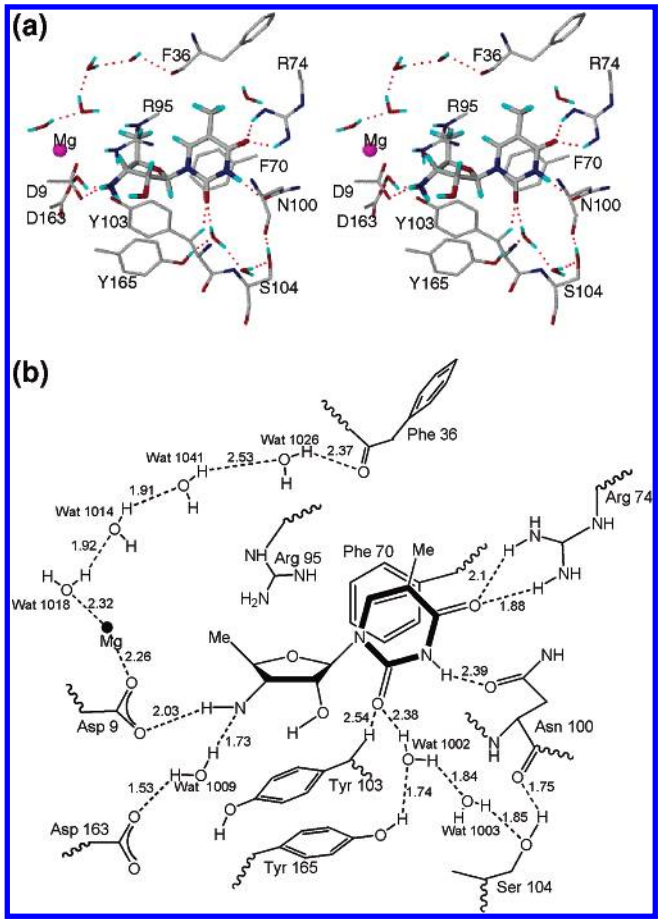


Figure 3. (a) Stereoview of the least-active molecule **14**−$TMPK_{mt}$ complex. (b) Schematic representation of interactions of the least-active molecule **14**−$TMPK_{mt}$ complex in the active site.

sites forming cooperative networks explains the important role of water in the binding affinities of the molecules.[43]

Pharmacophore. Ten pharmacophore models with a cost difference of 44.77 were generated. Table 2 gives the statistical parameters for all 10 models. Among these, the model with the lowest total cost value (78.68), lowest rmsd (0.66), and highest correlation coefficient (0.963) was chosen for further analysis. All of the training set molecules were predicted with an error less than 1 unit. The best model is further validated externally using a test set of 27 molecules, 77% of which were predicted with an error less than 1 unit. The actual and predicted activities of the training and test set molecules are given in Table 1.

The best pharmacophore model contains four features (Figure 4). The green colored contours represent three hydrogen-bond acceptors (HBAs) along with the directionality of the hydrogen bond, shown as an arrowhead (HBA_{1-3}), whereas the blue contour represents a HY feature. The most active, **38**, had a fitness score[44] of 7.36 when mapped to the pharmacophore (Figure 4a), whereas the least active, **14**, maps to a value of 5.05 (Figure 4b). In **38**, the HY feature corresponds to the Br atom at thymine C_5. Of the three HBAs, HBA_1 and HBA_2 correspond to the carbonyl O atoms of thymine C_4 and C_2, respectively, whereas HBA_3 corresponds to the hydroxyl oxygen at ribose $C_{3'}$. All four pharmacophore features are mapped in the training set molecules with pK_i values greater than 3.7. For molecules with lesser activity (**7**, **8**, **26**, **27**, **44**, and **46**), at least one

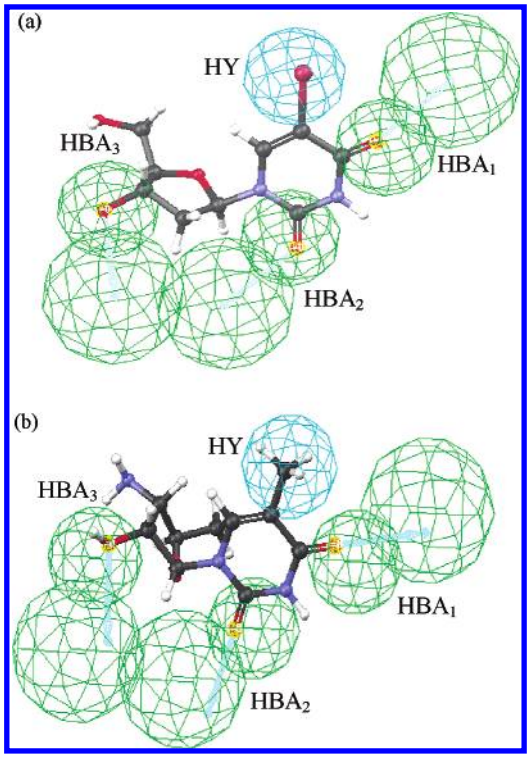


Figure 4. Pharmacophore mapping of the compounds: (a) most-active molecule, **38**, and (b) least-active molecule, **14**. The green contours represent the positioning of HBAs with their directionality indicated by arrowheads. Cyan contours represent the HY features.

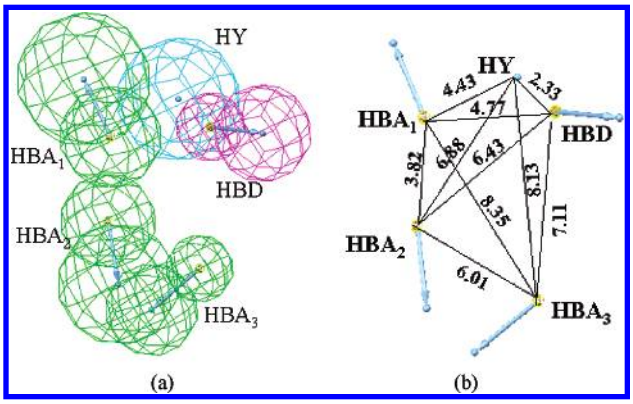


Figure 5. (a) Features of the composite pharmacophore model. HBAs are shown in green, HBDs in purple, and HY features in cyan. (b) Schematic representation of the five-point composite pharmacophore model showing their three-dimensional spatial relationship.

feature was missing. For example, **8**, **44**, and **46** miss the HY feature. Although all of the features are present in the least-active molecule, **14**, they do not fit the pharmacophore properly. In the test set, molecules **13** and **16** were predicted with an error greater than 1.5 units. Molecule **13** has an azide group at ribose $C_{3'}$ similar to **35**, **36**, and **37** (which have activity in the 4.5 pK_i range). Also, it has a hydroxyl at ribose $C_{2'}$. Because the pharmacophore does not take this into account, the activity is overestimated. In **16**, the spacing of the azide group with a methylene displaces the N atom beyond the HBA_3 contour. Thus, the activity is underestimated.

Database Screening. The pharmacophore model as developed so far gives a hopefully accurate idea of the

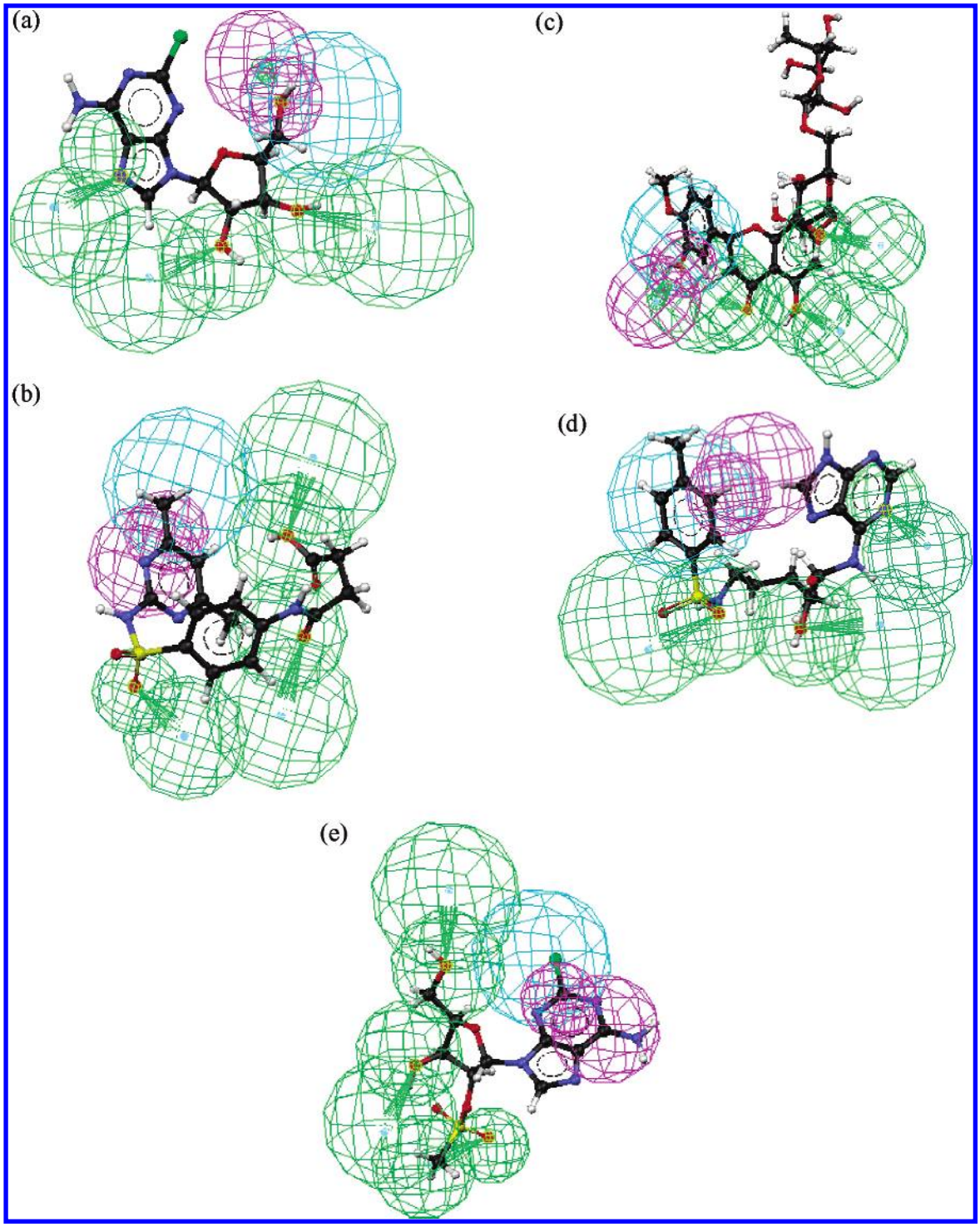


Figure 6. Pharmacophore mapping of representative hits from each of the five clusters.

necessary and sufficient molecular attributes required in a new lead. However, actual application of the model in VS requires that it be further validated. This validation and screening was carried out in three steps: (1) construction of a composite 3D pharmacophore model, (2) validation of this model by "spiking" it with known inhibitors, and (3) screening a large database to find potential novel leads.

Composite Model. The pharmacophoric features obtained were complementary to the active site residues and bound water molecules. The pharmacophore feature HBA_1 is complementary to the guanidino group of Arg74, whereas HBA_2 accepts a hydrogen bond from Wat1002 that, in turn, forms a hydrogen bond with the side-chain hydroxyl of Tyr165. HBA_3 accepts a hydrogen bond from Wat1009 and, in turn, forms a hydrogen bond with the side-chain carbonyl oxygen of Asp163; finally, the HY feature represents the cavity lined by the side chains of Arg95, Pro37, and Ser99. However, no feature corresponds to the N1 of the thymidine ring involved in a hydrogen bond with the carbonyl oxygen of Asn100. This is mainly due to the limitation of the hypogen-derived pharmacophore model to have a maximum number of seven points. (Whereas all of the pharmacophore features have one point, HBD and HBA features count as two points.) The hydrogen bond to Asn100 is one of the

critical interactions of TMP within the binding site in *M. tuberculosis* when compared with yeast, human, or *Escherichia coli* enzyme structures. This interaction was, therefore, seen to occur in the docking analysis, across all of the study molecules. Therefore, a composite pharmacophore model was built including this feature, as in Figure 5. Figure 5b gives the distance constraints between all of the features of the composite model.

Validation of the Composite Model. The composite model was validated by screening a database of compounds with known biological activities and calculating the enrichment factor (E)[30] using eq 4

$$E = \frac{H_a/H_t}{A/D} \tag{4}$$

where H_t = the number of hits retrieved, H_a = the number of actives in the hit list, A = the number of active molecules present in the database, and D = the total number of molecules in the database.

For this validation experiment, the MiniMaybridge database (containing 2000 molecules) was spiked with the 47 known inhibitors (A) used in this study. When this spiked database (containing 2047 molecules; D) was screened with the pharmacophore model, 110 molecules (H_t) were retrieved from the database as hits. Among these hits, 42 (H_a) molecules were from the 47 known actives. Thus, the *enrichment factor* (as per eq 4) was found to be 17.31, indicating that it is 17 times more probable to pick an active compound from the database than an inactive one.

Screening. The validated pharmacophore model was used to screen a large in-house database of 500 000 compounds (screen 1). About 5200 molecules were obtained as hits from our initial screen. To assess the drug-likeness of these hits, a second screen, incorporating Lipinski's rules, was used. A total of 1210 molecules was obtained as hits from this screen. To further increase the probability that a hit is a lead, a fitness score $\geq$ 3.00 was used as the third screen. All of this ensures that these hits include the bioactive conformation within a narrow energy threshold. A total of 186 molecules were obtained as hits from the third screen. These hits are comprised mostly of thymidine, benzopteridine phenylsulfonamide, and purine derivatives. To understand the chemical space of the final hits, a cluster analysis using a Tanimoto similarity index of fingerprints was carried out. All of the hit molecules were grouped into five clusters. Figure 6 gives the pharmacophore mapping of representative hit molecules from each cluster. Cluster 1 included purine and pyrimidine derivatives of ribose sugar, whereas sulfonamide derivatives constituted cluster 2. Flavanol derivatives formed cluster 3, and clusters 4 and 5 included one molecule each that are hybrid purine−sulfonamide structures.

CONCLUSIONS

A detailed docking analysis and pharmacophore modeling was carried out using $TMPK_{mt}$ inhibitors. Docking confirmed the role of weak interactions in promoting enzyme selectivity toward deoxyribonucleotides. It also highlighted the importance of water-mediated cooperative networks and weak hydrogen bonds to ligand binding affinities. Another interesting finding of this study is the role of halogen bonding and its stabilzation by the water-mediated cooperative hydrogen-bond network. With an appreciation of the functionalities involved in molecular recognition acquired from the above-mentioned methods, a composite pharmacophore model was developed and validated with a database containing known $TPMK_{mt}$ inhibitors. This composite model was used as a 3D query for the successful VS of a database of about 500 000 compounds to find new antitubercular leads.

ACKNOWLEDGMENT

G.R.D. thanks the CSIR (NMITLI) for their financial support of the project (Grant 5/258/6/2002-NMITLI). V.A., J.J., and M.R. thank the CSIR for fellowship support. G.R.D., V.A., J.J., and M.R. acknowledge the Centre for Modeling, Simulation, and Design (CMSD) set up in the University of Hyderabad under the UGC program for Universities with Potential for Excellence (UPE).

Supporting Information Available: Table of hits obtained from VS (Table 1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894.
(2) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.
(3) Xue, L.; Stahura, F. L.; Bajorath, J. Similarity search profiling reveals effects of fingerprint scaling in virtual screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2032−2039.
(4) Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. Virtual screening for SARS−CoV protease based on KZ7088 pharmacophore points. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1111−1122.
(5) Debnath, A. K. Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of *Mycobacterium avium* complex dihydrofolate reductase. *J. Med. Chem.* **2002**, *45*, 41−53.
(6) Srinivasan, J.; Castellino, A.; Bradley, E. K.; Eksterowicz, J. E.; Grootenhuis, P. D. J.; Putta, S.; Stanton, R. V. Evaluation of a novel shape-based computational filter for lead evolution application to thrombin inhibitors. *J. Med. Chem.* **2002**, *45*, 2494−2500.
(7) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(8) Schneider, G.; Böhm, H. J. Virtual screening and fast automated docking methods. *Drug Discovery Today* **2002**, *7*, 64−70.
(9) Krovat, E. M.; Langer, T. Impact of scoring functions on enrichment in docking-based virtual screening: an application study on renin inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1123−1129.
(10) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein−ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(11) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.
(12) Grüneberg, S.; Stubbs, M. T.; Klebe, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J. Med. Chem.* **2002**, *45*, 3588−3602.
(13) Steindl, T.; Langer, T. Influenza virus neuraminidase inhibitors: Generation and comparison of structure-based and common feature pharmacophore hypotheses and their application in virtual screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1849−1856.
(14) Stokstad, E. Infectious disease: Drug resistant TB on the rise. *Science* **2000**, *287*, 2391.
(15) Dye, C.; Williams, B. G.; Espinal, M. A.; Raviglione, M. C. Erasing the world's slow stain: Strategies to beat multidrug resistant tuberculosis. *Science* **2002**, *295*, 2042−2046.
(16) Haouz, A.; Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Herdewijn, P.; Calenbergh, S. V.; Delarue, M. Enzymatic and structural analysis of inhibitors designed against *Mycobacterium tuberculosis* thymidylate kinase. *J. Biol. Chem.* **2003**, *278*, 4963−4971.
(17) Lavie, A.; Ostermann, N.; Brundiers, R.; Goody, R. S.; Reinstein, J.; Konrad, M.; Schlichting, I. Structural basis for efficient phosphory-

lation of 3′-azidothymidine monophosphate by *Escherichia coli* thymidylate kinase. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14045−14050.
(18) Ostermann, N.; Segura-Peña, D.; Meier, C.; Veit, T.; Monnerjahn, C.; Konrad, M.; Lavie, A. Structures of human thymidylate kinase in complex with prodrugs: Implications for the structure-based design of novel compounds. *Biochemistry* **2003**, *42*, 2568−2577.
(19) Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Dugué, L.; Heyerick, A.; Keukeleire, D. D.; Pochet, S.; Busson, R.; Herdewijn, P.; Calenbergh, S. V. 3′-*C*-branched-chain-substituted nucleosides and nucleotides as potent inhibitors of *Mycobacterium tuberculosis* thymidine monophosphate kinase. *J. Med. Chem.* **2003**, *46*, 3811−3821.
(20) Pochet, S.; Dugué, L.; Labesse, G.; Delepierre, M.; Munier-Lehmann, H. Comparative study of purine and pyrimidine nucleoside analogues acting on the thymidylate kinases of *Mycobacterium tuberculosis* and of humans. *ChemBioChem* **2003**, *4*, 742−747.
(21) Sierra, L. D. L.; Munier-Lehmann, H.; Gilles, A. M.; Bârzu, O.; Delarue, M. X-ray structure of TMP kinase from *Mycobacterium tuberculosis* complexed with TMP at 1.95 Å resolution. *J. Mol. Biol.* **2001**, *311*, 87−100.
(22) Vanheusden, V.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Calenbergh, S. V. Synthesis and evaluation of thymidine-5′-O-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2695−2698.
(23) Vanheusden, V.; Rompaey, P. V.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Calenbergh, S. V. Thymidine and Thymidine-5′-O-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3045−3048.
(24) *Cerius*^{2}, molecular modeling program package; Accelrys: San Diego, CA.
(25) *InsightII*, molecular modeling program package; Accelrys: San Diego, CA, 1997.
(26) *GOLD*, version 2.0; Cambridge Crystallographic Data Centre: Cambridge, U. K.
(27) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(28) *Catalyst*, version 4.6; Accelrys: Burlington, MA.
(29) Smellie, A.; Teig, S. L.; Towbin, P. Poling: Promoting conformational coverage. *J. Comput. Chem.* **1995**, *16*, 171−187.
(30) Güner, O. F. *Pharmacophore: perception, development and use in drug design*; International University Line: La Jolla, CA, 2000.
(31) InterBioScreen homepage. http://www.ibscreen.com.
(32) ChemStar homepage. http://www.chemstar.ru.
(33) Molecular Diversity Preservation International homepage. http://www.mdpi.org.
(34) Key Organics Ltd homepage. http://www.keyorganics.ltd.uk.
(35) ASINEX−Intelligent Chemistry homepage. http://www.asinex.com/welcome.htm.
(36) http://www.apps1.niaid.nih.gov/struct_search/an/an_search.htm.
(37) Maybridge homepage. http://www.maybridge.com/html/home.htm
(38) The CSD search was restricted to entries with crystallographic $R < 0.10$ and no disorder. Also, metal complexes were excluded from the search.
(39) Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *J. Am. Chem. Soc.* **1996**, *118*, 3108−3116.
(40) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789−16794.
(41) Steinrauf, L. K.; Hamilton, J. A.; Braden, B. C.; Murrell, J. R.; Benson, M. D. X-ray crystal structure of the Ala-109 → Thr variant of human transthyretin which produces euthyroid hyperthyroxinemia. *J. Biol. Chem.* **1993**, *286*, 2425−2430.
(42) Jeffrey, G. A.; Saenger, W. *Hydrogen bonding in biological structures*; Springer-Verlag: Berlin, 1991.
(43) Sarkhel, S.; Desiraju, G. R. N−H···O, O−H···O and C−H···O hydrogen bonds in protein−ligand complexes: strong and weak interactions in molecular recognition. *Proteins* **2004**, *54*, 247−259.
(44) Fitness values indicate how well the features in the pharmacophore overlap with the chemical features in the ligand. These are normalized for the number of features in the pharmacophore model.

CI050064Z