JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Dynamics Based Pharmacophore Models for Screening Potential Inhibitors of Mycobacterial Cyclopropane Synthase

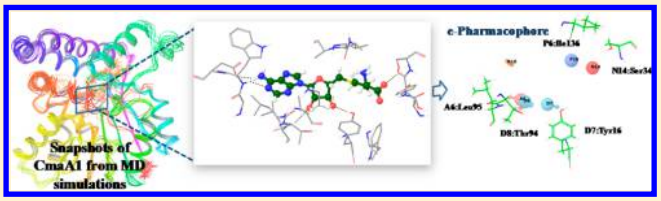
Chinmayee Choudhury,[†,‡] U. Deva Priyakumar,*[,†] and G. Narahari Sastry*[,‡]

[†]Centre for Computational Natural Sciences and Bioinformatics, International Institute of Information and Technology, Hyderabad 500032, India

[‡]Centre for Molecular Modeling, Indian Institute of Chemical Technology, Hyderabad 500007, India

Supporting Information

ABSTRACT: The therapeutic challenges in the treatment of tuberculosis demand multidisciplinary approaches for the identification of potential drug targets as well as fast and accurate techniques to screen huge chemical libraries. Mycobacterial cyclopropane synthase (CmaA1) has been shown to be essential for the survival of the bacteria due to its critical role in the synthesis of mycolic acids. The present study proposes pharmacophore models based on the structure of CmaA1 taking into account its various states in the cyclopropanation process, and their dynamic nature as assessed using molecular dynamics (MD) simulations. The qualities of these pharmacophore models were validated by mapping 23 molecules that have been previously reported to exhibit inhibitory activities on CmaA1. Additionally, 1398 compounds that have been shown to be inactive for tuberculosis were collected from the ChEMBL database and were screened against the models for validation. The models were further validated by comparing the results from pharmacophore mapping with the results obtained from docking these molecules with the respective protein structures. The best models are suggested by validating all the models based on their screening abilities and by comparing with docking results. The models generated from the MD trajectories were found to perform better than the one generated based on the crystal structure demonstrating the importance of incorporating receptor flexibility in drug design.



INTRODUCTION

Tuberculosis (TB), caused by the pathogen *Mycobacterium tuberculosis* (*M. Tb.*) remains one of the deadly human diseases in spite of significant efforts to eradicate it. According to the World Health Organization (WHO), increase in the global incidence of this disease is attributed to several factors including inadequate or poorly administered treatment regimens, insufficient supply or quality of anti-TB medicines, transmission of drug resistant strains, and coinfection with HIV-AIDS.[1] Describing the situation as a "global health emergency", WHO estimates that nearly half a million multidrug-resistant tuberculosis (MDR-TB) cases emerge worldwide each year.[1] Hence, new and efficient drugs are a desperate need of the hour to treat TB, which necessitate identification of potential drug targets.[2−5] In recent decades, computational methods have become mainstream in understanding the ligand−receptor interactions as well as for in silico screening of huge chemical libraries providing a fast and less expensive alternative to the traditional high throughput screening.[6−10] Also, rapid developments in high throughput NMR spectroscopy and X-ray crystallography[11,12] have paved the path of rapid structure determination of the targets and have accelerated the use of structure based virtual screening methods by many folds.[13,14] Incorporating receptor flexibility in docking calculations in an efficient way is a major challenge, and few methods have been proposed to address this. Few of the early methods included "soft-docking" by Jiang et al.[15] where the penalty for van der Waals clashes between the receptor and ligand atoms was reduced followed by other researchers who significantly contributed toward development and improvements of the soft docking methods.[16−18] Leach developed an algorithm that considers the conformational flexibilities of amino acid side chains and ligands using a rotamer library and identifies the combinations of conformations of side chains and the ligand with lowest energies.[19] Some examples include the studies of Murray et al.[20] on thrombin, thermolysin, and neuraminidase, studies by Bouzida et al.[21] and Erickson et al.[22] on HIV Protease, Cavasotto and Abagyan's flexible-ligand-grid receptor docking experiments on 33 crystal structures of 4 protein kinase subfamilies,[23] Schapira and Abagyan's study on virtual screening against nuclear receptors,[24] and a docking study by Daeyaert et al.[25] using the crystal structures of HIV reverse transcriptase complexed with different non-nucleoside inhibitors. Representation of the flexibility of receptors through collective degrees of freedom also became popular in the early 1990s.[26,27] Use of multiple protein conformations offers the ligand to select from an ensemble of pre-existing partially fitting protein conformations which represent the highly populated low-energy states of a receptor also founded a strong basis for incorporating protein flexibility in structure based drug design.[28−30]

Received: December 13, 2014
Published: March 9, 2015

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

MD simulation has emerged as a powerful and efficient method to address the receptor flexibility issues in structure based drug design. Carlson et al.[31] developed a "dynamic" pharmacophore model of HIV integrase from an ensemble of MD snapshots and identified two new inhibitors screening the Available Chemical Database (ACD). They also validated this method by generating a dynamic pharmacophore model from the apo form of HIV protease, which could discriminate between known HIV-1 protease inhibitors and other drug-like molecules.[32] Another approach by Moitessier et al.[33] involves prediction of binding modes by analyzing the dynamic pharmacophore model and orientation of the pharmacophore points. Several studies have been carried out by Carlson and co-workers, which demonstrated the successful incorporation of protein flexibility in structure based drug design applied over a range of pharmacologically relevant targets.[34−39] Mycobacterial CmaA1 is an essential drug target, which is responsible for cis-cyclopropanation at the distal position of α-mycolates. In our earlier work, we had studied the structural, energetic and dynamic properties of CmaA1 by performing 40 ns MD simulations on each of the five model systems of CmaA1 representing various stages of cyclopropanation process.[40] In this study, we have made an attempt to incorporate dynamic nature of the active site of CmaA1 in drug design by generating structure based pharmacophore models from the snapshots taken from the MD trajectories. These models have been validated by examining their efficiencies to screen previously known CmaA1 inhibitors.[40] The results from the models were also compared with docking calculations done on multiple conformations of the protein structure.

METHODOLOGY

Model Systems. The following five model systems of CmaA1 were considered for generation of dynamics based pharmacophore models. **E-SAM**: Holo form of CmaA1 with SAM in the cofactor binding site (CBS). **E-SAM-S**: Holo form with SAM in the CBS and the substrate in the acetyl substrate binding site (ASBS). **E-SAHC-P**: Holo form with SAHC in the CBS and the cyclopropanated product in the ASBS. **E-SAHC**: Holo form with SAHC in the CBS. **E-SAHC-D**: Holo form with SAHC in the CBS and a CmaA1 inhibitor DDDMAB in the ASBS. The first four systems were taken from our previous study,[41] and the fifth model (**E-SAHC-D**) was based on the MD simulations done on the crystal structure of the holo state of CmaA1 structure (PDB ID: 1KPH[42]).

Molecular Dynamics (MD) Simulations. The methodology for MD simulation with the CHARMM program[43] on the system **E-SAHC-D** was the same as for rest of the four systems. The individual components of each system were generated using CHARMM GUI[44] using the CGenFF force field.[45] Nonphysical contacts were removed by minimizing the whole system using a 1000-step adopted basis Newton−Raphson (ABNR) minimization. The system was then solvated in a TIP3P water box of 80 Å × 75 Å × 75 Å dimension and appropriate numbers of counterions were added for neutralization. A 1 ns equilibration run was used in the NVT ensemble with periodic boundary conditions after 500 steps of steepest descent (SD) and 500 steps of ABNR energy minimizations. The solute molecules were restrained by applying a constant harmonic force of 5 kcal/mol during the initial minimization and equilibration. The systems was then subjected to 40 ns MD simulation in the NPT ensemble at 298 K with leapfrog integrator and a time step of 0.002 ps. Langevin piston algorithm, SHAKE, and particle mesh Ewald (PME) were used to perform pressure control, constraint covalent bonds involving hydrogen atoms and to treat long-range interactions, respectively. The coordinates were saved every 5 ps for analysis.

Generation of Structure Based Pharmacophore Models. Forty snapshots were collected at every 5 ns interval from the MD trajectories of the five model systems of CmaA1. Including the static crystal structure 1KPH, a total of 41 structures were thus considered to construct the structure based e-pharmacophore models. Glide energy grids were generated for each snapshot to define the active site as a cubical box of 12 Å × 12 Å × 12 Å around the cofactors and then the interactions between the cofactor and protein was assessed by using "Score in place" mode of the Glide, Schrödinger molecular modeling suite, with the option to output Glide extra precision (XP) descriptor information.[46] Default settings were used for the scoring. The resulting protein−cofactor complexes along with the XP energy terms were then subjected to the e-Pharmacophore[47] generation tool of Schrodinger to generate energy based pharmacophore models. Figure 1 schematically shows the generation and selection of the e-Pharmacophore models as filters for virtual screening.

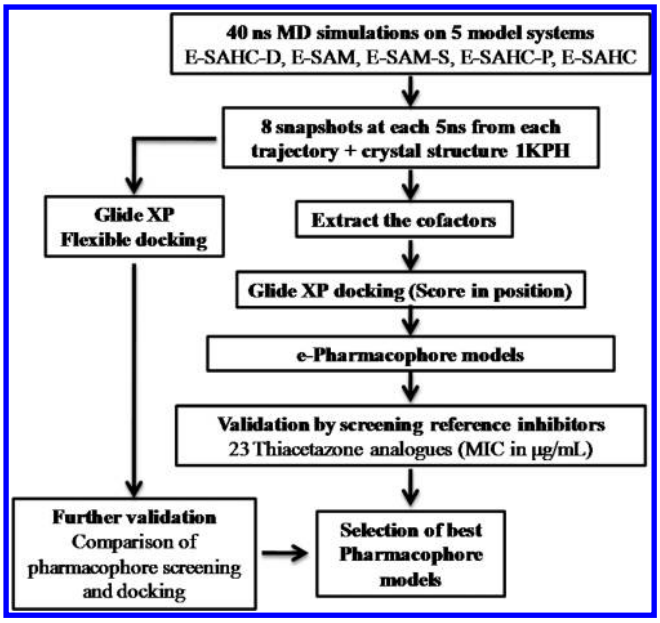
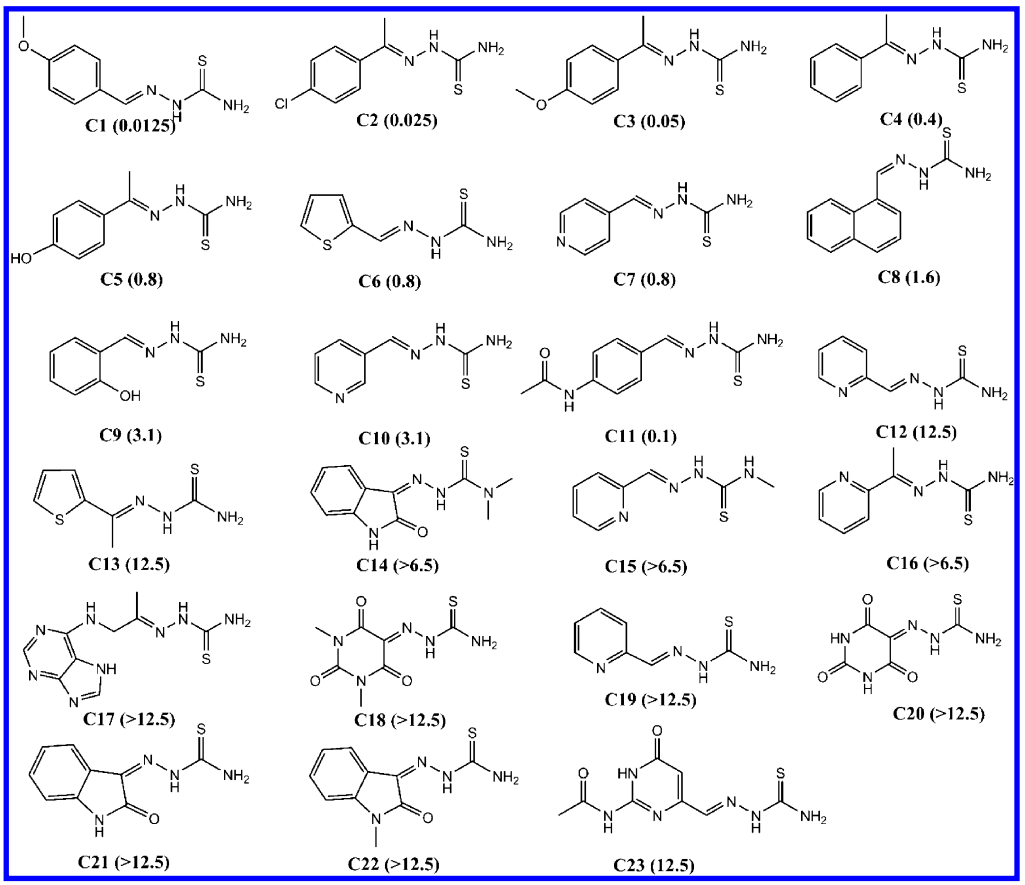


Figure 1. Schematic representation of the generation of various types of pharmacophore models as filters for virtual screening.

Pharmacophore Screening and Docking. A set of 23 CmaA1 inhibitors with reported MIC values ranging from 0.0125 to 12.5 μg/mL[40] (Scheme 1) were used for verifying the abilities of all the generated structure based e-Pharmacophore models. All these compounds were energy minimized using the default parameters of LigPrep[48] module of the Schrodinger Suite. Five best conformers were chosen for each compound. The "Advanced Pharmacophore Screening" option, Phase module of the Schrodinger Suite,[49] was used with an option to generate five conformations per rotatable bond and maximum number of conformations per compound were kept to be 100. A rapid sampling was used for screening and the default option for skipping structures with more than 15 rotatable bonds was used. The minimum number of sites the

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

Scheme 1. Compounds Used for Validation of the Performance of the Pharmacophore Models to Screen Active Inhibitors[a]



[a]The compound name and the MIC values (μg/mL) are given below each compound.[40]

molecule must match was assigned to be four for all the models. Among many conformers of a ligand, the one with the best fitness score (S) given by the following equation[49] was retained for each compound.

$$S = W_{site}(1 - S_{align}/C_{align}) + W_{vec}S_{vec} + W_{vol}S_{vol} + W_{ivol}S_{ivol}$$

where, S_{align} is the alignment score, i.e., RMS deviation between the site point positions in the matching conformation and the site point positions in the hypothesis, C_{align} is the alignment cutoff, W_{site} is weight of site score $(1 - S_{align}/C_{align})$, S_{vec} is the vector score, i.e., average cosine between vector features in the matching conformation and the vector features in the reference conformation, W_{vec} is the weight of vector score, $S_{vol}(V_{common}/V_{total})$ is the volume score, i.e., ratio of the common volume occupied by the matching conformer and the reference conformer, to the total volume (the volume occupied by both). W_{vol} is the weight of volume score, S_{ivol} is the included volume score, i.e., ratio of the volume overlap between the matching conformer and the included volumes (if present) to the total included volume. Volumes were computed using van der Waals models of all atoms except nonpolar hydrogens, and W_{ivol} is the weight of volume score. C_{align}, W_{site}, W_{vec}, W_{vol}, and W_{ivol} are user-adjustable parameters, with default values of 1.20, 1.00, 1.00, 1.00, and 0.0, respectively. To verify if the models screen any inactive compounds, 2050 compounds reported to be inactive against *M. Tb.* were collected. 1398 compounds (Table S1) were found to be within the molecular weight range of 180−400 and consisting 12 to 27 heavy atoms (similar to that of the 23 active compounds, SAM and SAHC). These 1398 compounds were then screened against all e-Pharmacophore models using the same criteria to check if these models screen any inactive compounds.

For further validation, all the 23 reference inhibitors were docked to the active sites of each of the snapshots and the crystal structures. The compounds were first subjected to Glide SP[46] module to generate five best poses for every compound. The Glide SP generated poses were further subjected to XP docking and the poses with the highest score for all the compounds were retained which were ranked based on the docking score for each snapshot. The top scoring compounds from docking and the hits screened by the corresponding structure based e-Pharmacophore models were compared.

RESULTS AND DISCUSSION

Proteins being very flexible, only some of the conformations in physiological conditions are relevant for ligand binding. Hence, choosing the right conformations of binding pockets of receptors for structure based virtual screening processes is crucial for identifying suitable lead compounds. In our previous study, we observed significant conformational changes in the binding sites of CmaA1 with respect to the different stages of cyclopropanation. Given the possibility of diverse conformational states, considering receptor flexibility in this case seems to be crucial for structure based drug design. Many hypotheses also suggest that the binding site of a protein exists in an equilibrium ensemble of different conformations of similar

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

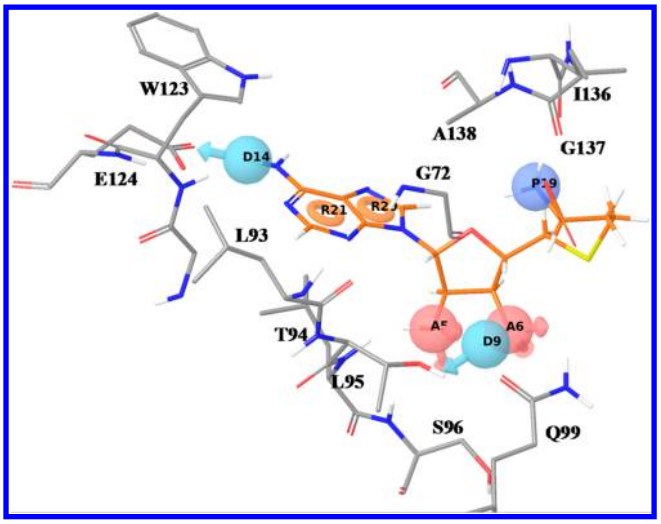


Figure 2. e-Pharmacophore model generated from the crystal structure of CmaA1 (1KPH) along with the associated active site residues.

energy, and ligands can choose the most favorable conformer from this ensemble with which to bind.[50,51] Hence, we collected the snapshots of holo CmaA1 structure bound to different cofactors/substrate/product from the MD simulation trajectories representing all the possible natural states of the target and exploring all types of interactions of the active site residues with the natural ligands SAM and SAHC. The crystal structure (1KPH) was also considered for generation of pharmacophore models as we were interested in comparing the predictive abilities of the pharmacophore models generated from the MD snapshots and the static crystal structure. e-Pharmacophore models were generated based on the interaction of SAM/SAHC with the active site residues of CmaA1 from each of the structures considered for the study (the MD snapshots and the static crystal structure). e-Pharmacophore is a new approach, which generates structure based pharmacophore models using the energetic binding terms from Glide XP and is a relatively fast method for screening. The protein−ligand interactions are reasonably well characterized in e-Pharmacophore by the XP terms, and excluded volumes corresponding to regions occupied by the receptor are also allowed. The models consist of six different chemical features, viz., hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic sites (H), negative ionizable sites (N), positive ionizable sites (P), and aromatic rings (R), and the maximum number of features per model was assigned as nine.[52] Hydrogen bond donors were represented as projected points, located at the corresponding hydrogen bond acceptor positions in the binding site. Projected points allow the possibility for structurally dissimilar active compounds to form hydrogen bonds to the same location, regardless of their point of origin and directionality. Each pharmacophore feature site is first assigned an energetic value equal to the sum of the Glide XP contributions of the atoms comprising the site, allowing sites to be quantified and ranked on the basis of the energetic terms. Glide XP descriptors include terms for hydrophobic enclosure, hydrophobically packed correlated

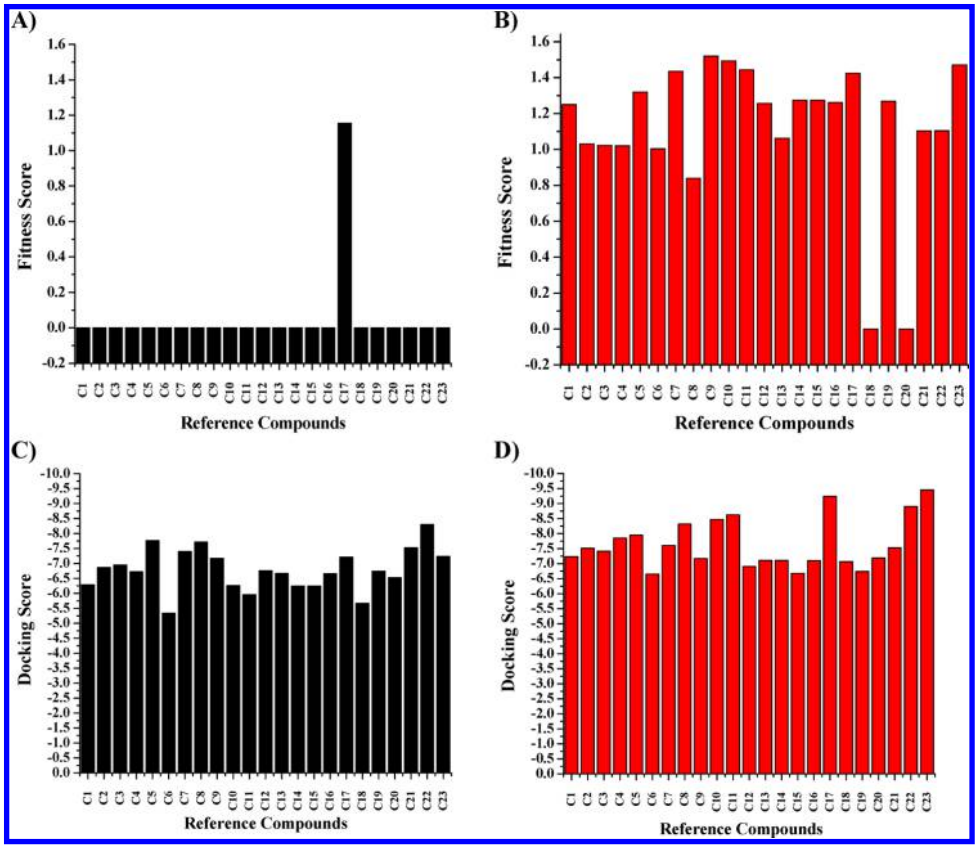


Figure 3. Plots showing the pharmacophore fitness and docking scores of the reference compounds with the crystal structure and the MD snapshots. Among 40 MD snapshots, the highest score for each of the compound have been shown here. The black columns indicate the scores with the crystal structure while the red columns indicate the scores with the MD snapshots.

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

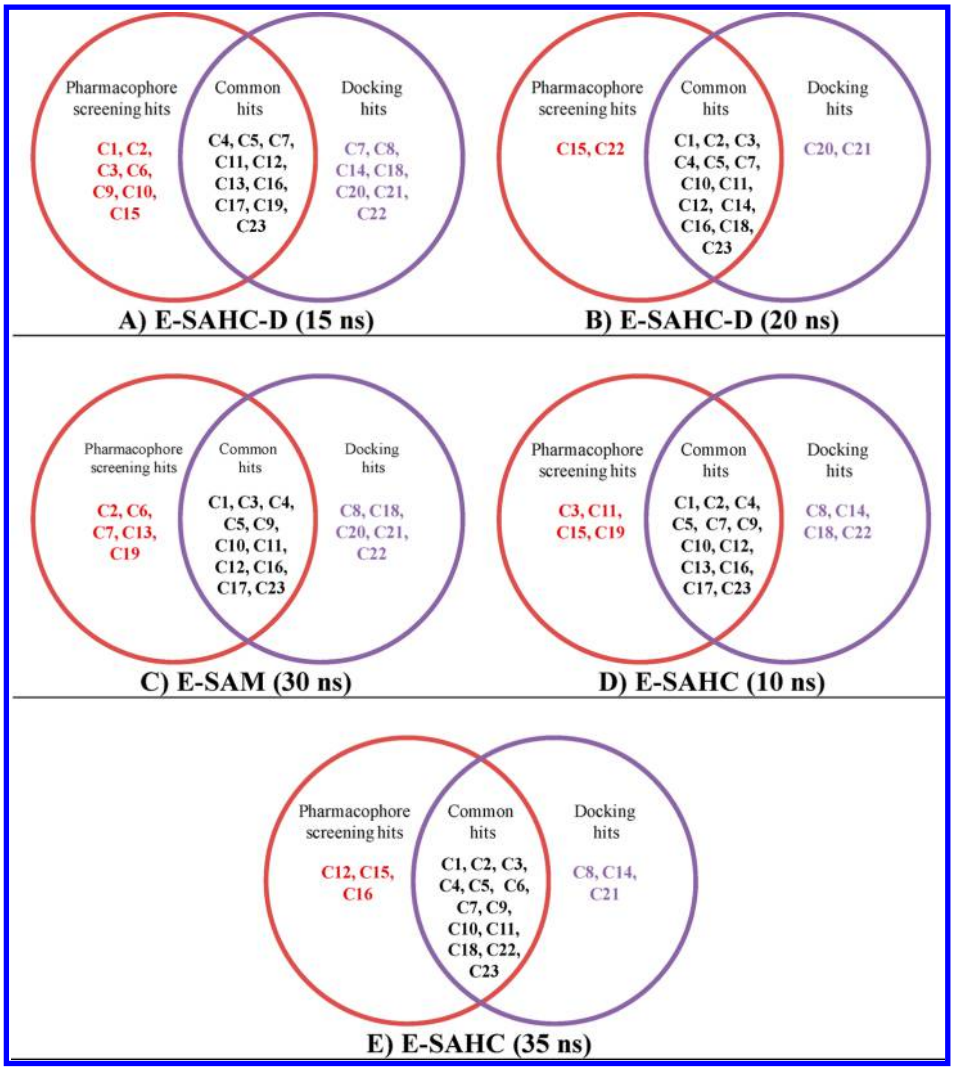


Figure 4. Compounds screened by the best 5 e-Pharmacophore models and docking with the respective snapshots.

hydrogen bonds, electrostatic rewards, $\pi-\pi$ stacking, cation$-\pi$, and other interactions. ChemScore hydrogen bonding and lipophilic atom pair interaction terms are included while the Glide XP terms for hydrogen bonding and hydrophobic enclosure are zero.

Comparison of e-Pharmacophore Models Generated from Different Model Systems. The e-Pharmacophore model generated from the static crystal structure has seven features (Figure 2). The A and D features near the sugar part of SAHC are formed as a result of the H-bonding between the residues T94, L95, and Q99. The terminal $-NH_3^+$ group interacts with G72, I136, and G137 resulting in the P feature. The $-NH_2$ group attached to the C6 atom of the adenine moiety of SAHC with E124 results in formation of a donor feature. Two R features are located near the aromatic ring of the adenine part of SAHC. The presence and the absence of the cofactor and or the ligand induce several conformational changes and the differences among the pharmacophore models due to this effect are discussed below.

Details of all the pharmacophore models are given in Table S2. Overall many similarities were found among the models generated from the trajectories of different model systems, the most common being the R features which occur near the F142 residue. The other common features found are the D and A features near the residues G72, T94, Y16, L95, S96, and Q99 as a result of H-bonds made by the sugar moiety of SAM/SAHC. The P and N features are found mostly near the residues I136, Y33, and S34, respectively, which show electrostatic interaction with the polar terminal part of the cofactors. In the system **E-SAM**, SAM has a different conformation than the other systems and interacts with different residues. So, the spatial locations of the features of models generated from the system **E-SAM** were found to be strikingly different than those of the others. H-bonding with the residues G72, S96, and Q99 with the sugar part of SAM was not found in this system as **E-SAHC-D**. Due to the conformational differences between SAM and SAHC, G72 interacts with the terminal $-NH_3^+$ group of SAM rather than the sugar part as in **E-SAHC-D**. Similarly, in the **E-SAM-S** system, the two $-$OH groups of the sugar part of SAM mostly make H-bonds with Q99 as H-bond acceptors and with Y16 as H-bond donors in some snapshots contributing toward the A and D features. Initially the terminal $-NH3^+$ and $-$COOH groups make H-bond and salt bridge interactions with G72 and S34, respectively, but after 10 ns, there is a large conformational change in SAM and new interactions were formed with D70 and T78 respectively in all snapshots giving rise to the P and N features. It was observed that the P features have higher scores in the **E-SAM** and **E-SAM-S** systems, while the A and D

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

Table 1. Features of the Selected Five Best e-Pharmacophore Models

feature label	type	score	X	Y	Z	source
			(A) **E-SAHC-D** (15 ns)			
D8	D	−1.61	−3.72	−6.27	4.90	HBond
A5	A	−1.60	−3.06	−6.97	4.84	HBond
A6	A	−1.52	−1.28	−4.83	5.28	HBond
R21	R	−0.97	−5.27	−8.56	0.97	RingChemscoreHphobe
N18	N	−0.73	−0.64	2.56	5.19	HBond
D9	D	−0.59	−2.21	−4.77	5.48	HBond
R20	R	−0.59	−3.20	−8.26	0.77	RingChemscoreHphobe
D13	D	−0.27	−6.10	−11.18	−0.95	HBond
			(B) **E-SAHC-D** (20 ns)			
A2	A	−1.95	−6.27	−9.45	0.03	PhobEnHB+HBond
A5	A	−1.60	−2.90	−7.40	4.16	HBond
D9	D	−1.60	−2.27	−4.91	4.67	none
R21	R	−0.92	−5.11	−8.72	0.31	RingChemscoreHphobe
D8	D	−0.83	−3.43	−6.63	4.36	HBond
N18	N	−0.56	−0.46	2.23	5.08	HBond
R20	R	−0.56	−3.01	−8.56	0.24	RingChemscoreHphobe
A6	A	−0.25	−1.34	−5.11	4.81	HBond
D14	D	−0.25	−5.88	−11.64	−0.96	HBond
			(C) **E-SAM** (30 ns)			
P16	P	−2.61	−4.07	0.08	3.59	Electro
A6	A	−1.60	−3.04	−8.72	5.56	HBond
D7	D	−1.60	−0.37	−6.73	6.50	HBond
R18	R	−0.56	−2.76	−9.49	0.46	RingChemscoreHphobe
D8	D	−0.47	−3.35	−8.00	6.09	HBond
N14	N	−0.33	−1.64	1.33	4.62	none+HBond
			(D) **E-SAHC** (10 ns)			
A2	A	−2.20	−6.81	−9.73	−0.11	PhobEnHB+HBond
A5	A	−1.60	−3.23	−8.70	5.46	HBond
D8	D	−1.60	−3.79	−8.31	6.14	HBond
D9	D	−1.60	−3.65	−5.84	6.75	HBond
A6	A	−1.26	−3.77	−5.89	5.80	HBond
D14	D	−0.34	−6.15	−11.90	−1.25	None
N18	N	−0.25	0.44	−0.01	4.78	none+HBond
R21	R	−0.89	−5.92	−9.20	0.76	RingChemscoreHphobe
			(E) **E-SAHC** (35 ns)			
A5	A	−1.60	−3.36	−8.55	4.48	None
D9	D	−1.60	−2.38	−5.80	6.31	HBond
A2	A	−1.50	−6.84	−9.70	−0.29	PhobEnHB
D8	D	−1.46	−3.70	−8.05	5.23	HBond
N18	N	−0.94	−0.70	1.68	4.84	HBond
R21	R	−0.87	−5.82	−9.13	0.41	RingChemscoreHphobe

features have higher scores in the **E-SAHC-D**, **E-SAHC-P** and **E-SAHC** systems. The models generated from the crystal structure do not have an N feature whereas the ones generated from the **E-SAHC-D** system do not have P features. Such differences among the pharmacophore models further illustrate the major variations in the conformational states of the residues that form the binding site, and hence the importance of accounting for the flexible nature of the protein.

Screening CmaA1 Inhibitors by e-Pharmacophore and Docking. The predictive abilities of the generated models were verified by screening a set of 23 reference compounds showing CmaA1 inhibitory activities in the range from 0.0125 to 12.5 μg/mL. These are antitubercular drug thiacetazone and its clinical analogues that are shown to cause significant loss of cyclopropanation in various mycobacterial strains.[40] The reversal of their effect on cyclopropanation upon overexpression of the cyclopropane synthase enzymes has provided evidence of direct binding of these compounds to cyclopropane synthase.[40] The pharmacophore fitness score was obtained for each of the compound against each model and hits were fetched in order of decreasing fitness. The fitness score is a linear combination of the site and vector alignment scores and the volume score that measures how well the matching pharmacophore site points align to those of the hypothesis, how well the matching vector features (acceptors, donors, aromatic rings) overlay those of the hypothesis, and how well the matching conformation superimposes, in an overall sense, with the reference ligand conformation.

The reference compounds were also docked to the active sites of the MD snapshots and the crystal structure to analyze the interactions made by the reference compounds with the active site residues of CmaA1 and compare these interactions with the pharmacophoric features of the corresponding e-Pharmacophore models matched by the same compounds. The

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

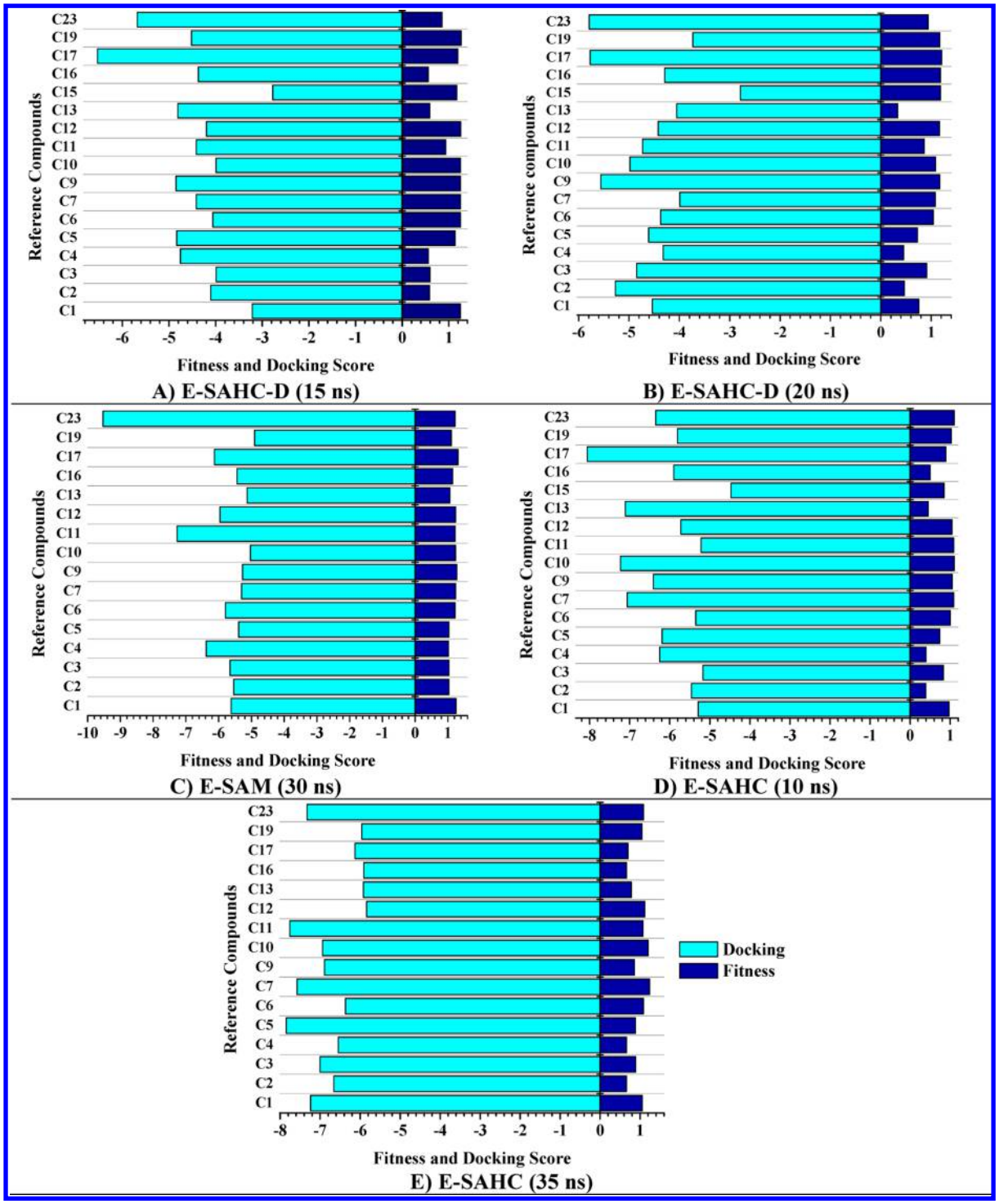


Figure 5. Pharmacophore fitness score and XP docking scores of the reference compounds with the models and the respective snapshots. The cyan bars represent the docking scores while the dark blue ones represent the pharmacophore fitness scores.

pharmacophore fitness score and the XP docking scores were compared for each molecule with each pharmacophore model and the corresponding snapshot. The pharmacophore fitness scores and docking scores for each of the reference compounds with all the 41 pharmacophore models/snapshots are given in Tables S3 and S4.

The models generated from the MD snapshots could screen up to 17 out of 23 reference compounds while the model generated from the crystal structure could screen only one compound. It was found that for a given compound the highest fitness score with the e-Pharmacophore models obtained from the MD snapshots was higher than that obtained from the crystal structure. The docking scores of a given reference compound with the crystal structure were mostly found to be lower than that with the MD snapshots. This shows the advantage of sampling many conformations of the binding site from the MD studies in screening. Figure 3 shows the pharmacophore fitness and docking scores for each reference compound with the models generated from the crystal structure and the MD snapshots. Table S2 shows the details of all

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

Table 2. Comparison of Number of Inhibitors and Noninhibitors Screened by the Selected Models And Ranges of Their Fitness Scores

model	% of inhibitors screened	fitness score range	% of noninhibitors screened	fitness score range
E-SAHC-D (15 ns)	74	0.56−1.56	4.5	0.4−1.06
E-SAHC-D (20 ns)	74	0.46−1.41	5.4	0.32−0.96
E-SAM (30 ns)	70	1.02-.1.51	0	-
E-SAHC (10 ns)	74	0.44−1.70	5.1	0.33−0.87
E-SAHC (35 ns)	70	0.66−1.64	0.5	0.38−0.91

pharmacophore models generated from the static crystal structure and snapshots taken at every 5 ns of MD trajectories of five model systems of CmaA1, the compounds screened by them and the common compounds that are matched by a model and also in the top scoring list upon docking with the corresponding snapshot.

It was also observed that at least one snapshot from each trajectory gave a model that could screen 13−17 out of 23 active compounds. At the same time models generated based on certain snapshots from each of the five trajectories screened none of the active compounds. This is in agreement with the hypothesis that ligand binding is an event where the ligand selects the most suitable binding site conformation from an ensemble of pre-existing partially fitting receptor conformations, as stated by Bosshard.[53,54] This concept of pre-existing receptor conformational ensembles offers a strong support to the incorporation of receptor flexibility through multiple receptor conformations generated by MD simulations.[50] As the model systems considered in this study are the representative conformations of the highly populated low-energy states of CmaA1 binding site when bound to its natural ligands (**E-SAM**, **E-SAM-S**, **E-SAHC-P**, **E-SAHC**) at each step of the cyclopropanation reaction, as well as when bound to an inhibitor (**E-SAHC-D**), the inhibitors could recognize one of

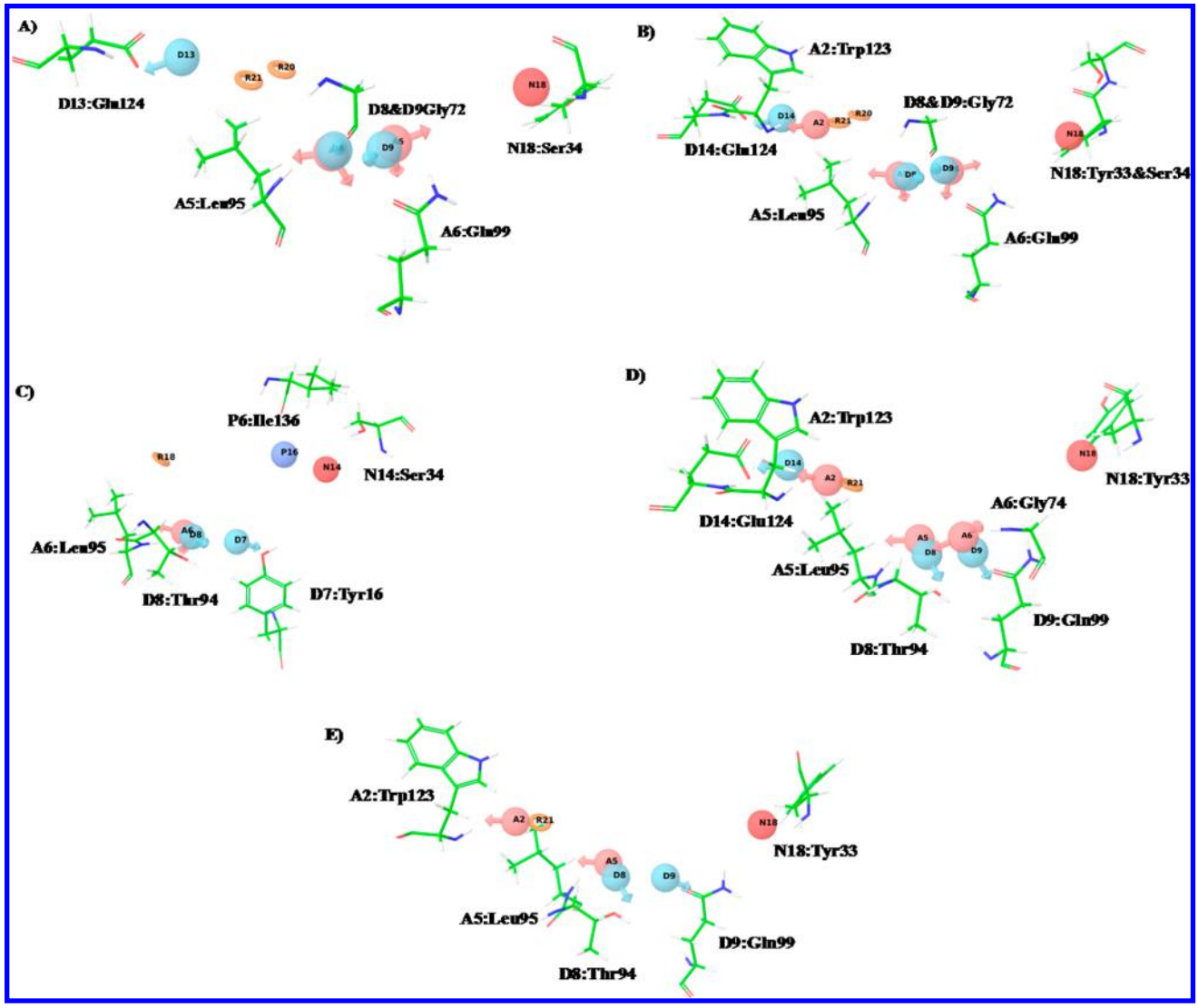


Figure 6. Selected e-Pharmacophore models with the active site residues associated with the pharmacophore features. (A) **E-SAHC-D** (15 ns), (B) **E-SAHC-D** (20 ns), (C) **E-SAM** (30 ns), (D) **E-SAHC** (10 ns), and (E) **E-SAHC** (35 ns). Color codes for the pharmacophoric features are as follows: cyan D, pink A, red N, blue P, green H, and orange R. Same color code for the features is followed for all the other figures.

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

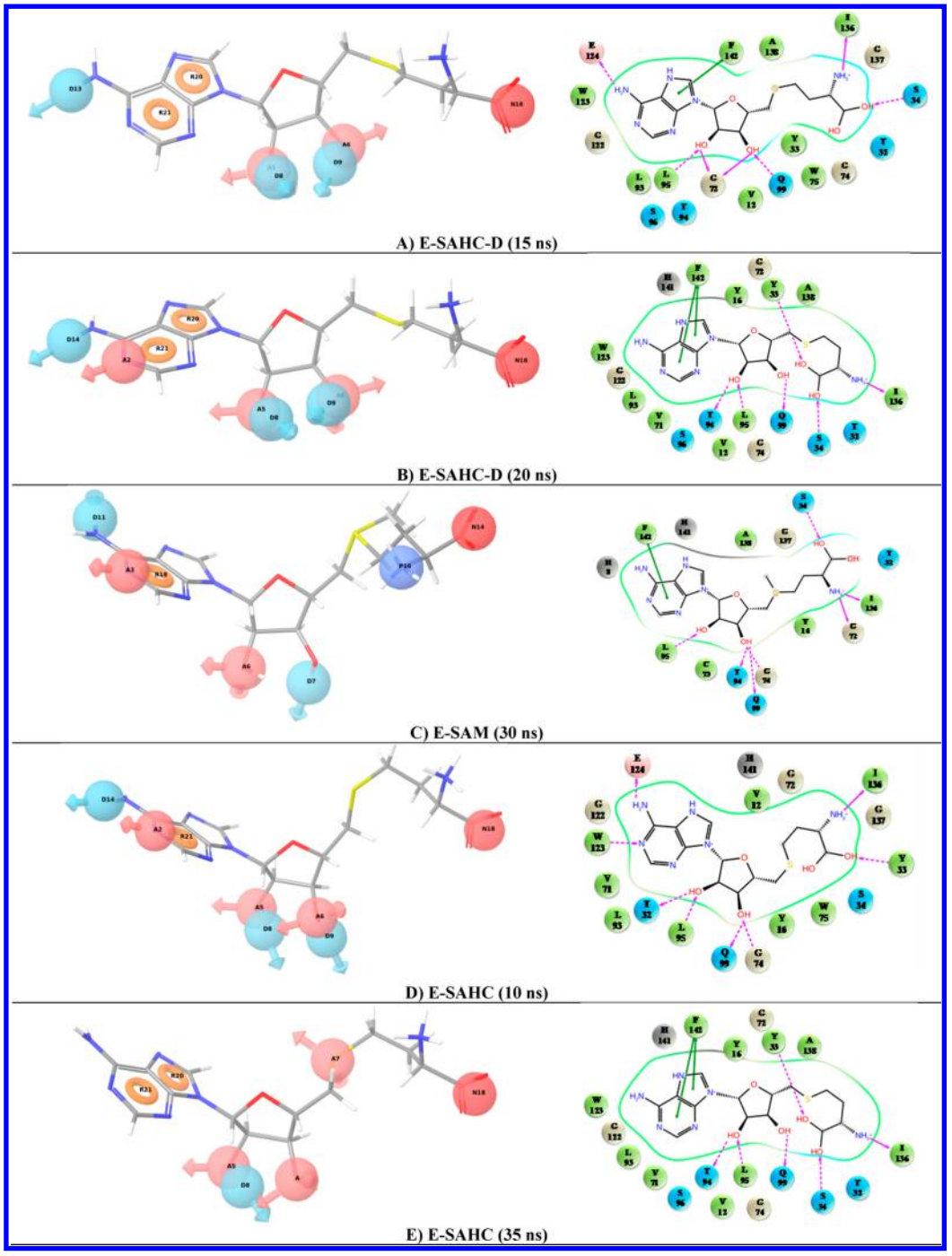


Figure 7. Best five e-Pharmacophore models mapped to SAM/SAHC and the interactions of SAM/SAHC with the active site residues which give rise to the pharmacophoric features.

those representative conformation at each step of cyclopropanation that best suits them for binding. A single static crystal structure of CmaA1 does not appropriately represent all such possible states and, hence, does not account for binding of all the ligands considered.

The generated e-Pharmacophore models were also tested for their ability to differentiate between the inhibitors and noninhibitors of similar size and molecular weight. For this purpose, 1398 noninhibitors of *M. Tb.* were collected from the ChEMBL database. Upon screening with the e-Pharmacophore models using the same criteria, it was observed that none of the models could screen more than 180 non inhibitors out of 1398. Also, the ranges of the fitness scores for the noninhibitors (Table S2) were found to be lower than those of the inhibitors. The maximum fitness scores for noninhibitors screened by each of the models were mostly found to be less than 1, and mostly lower than the active compounds.

Comparing the screening and docking results, we found that the reference compound C17 and C23 were screened by 24 and 26 out of 41 pharmacophore models respectively, and interestingly, these two compounds showed the highest scores viz., −9.2 and −9.5 with the snapshots **E-SAM** at 30 ns and **E-SAHC-P** 10 ns. The correlation among the pharmacophore fitness and docking score is insignificant because it is well-

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

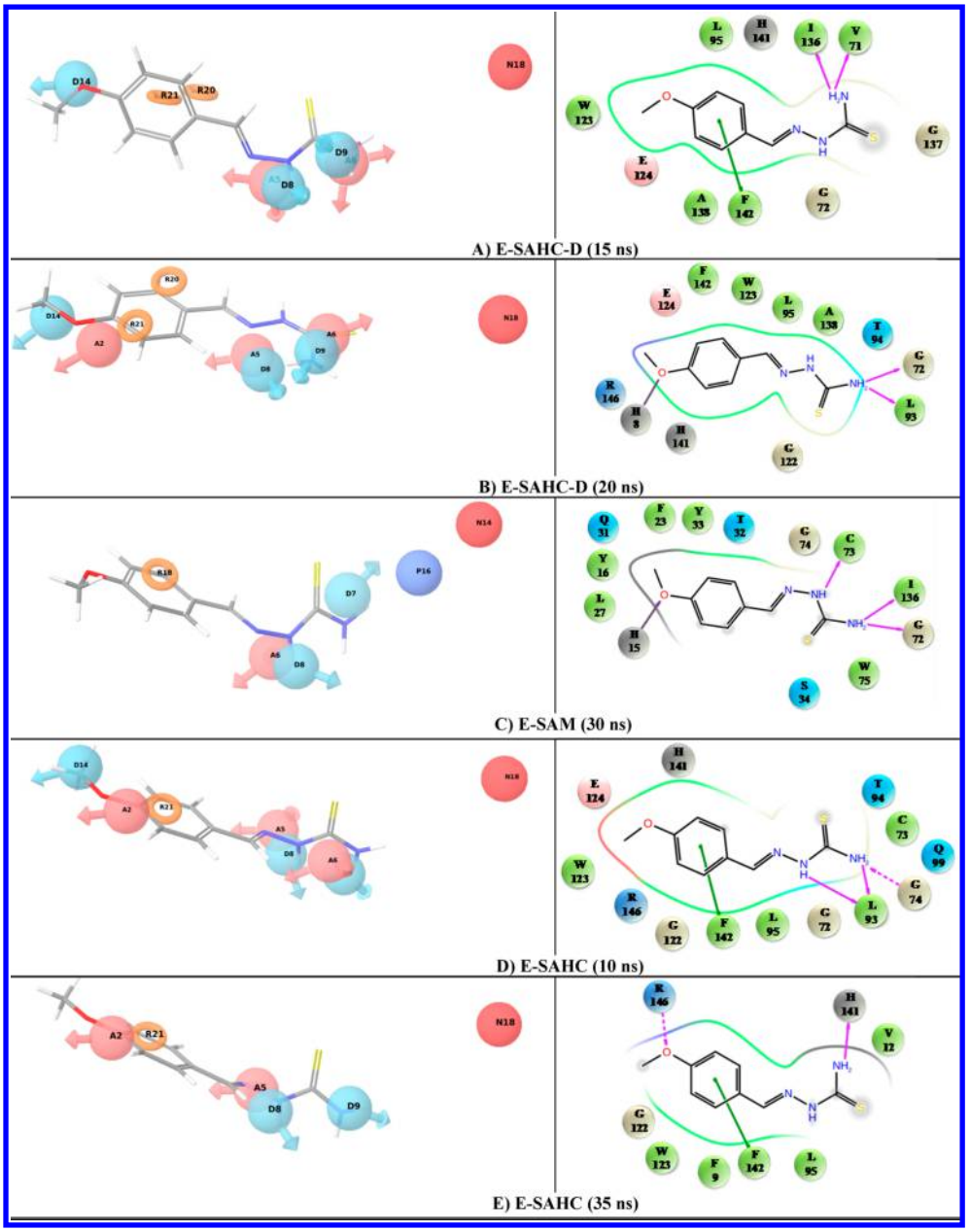


Figure 8. Most active reference compound C1 mapped with the best five e-Pharmacophore models and the docked pose and interaction of the compound with the active site residues of the corresponding snapshots.

known that both these methods use many approximations while calculating the scores. Also the activity range of the reference compounds was also not very wide, and we consider all of them to be active. Hence, a lower correlation between these two scores and also with the activity is expected. Considering the fact that both pharmacophore screening and docking are recognized as integral parts of drug design, each method having its own strengths and weaknesses, we tried to identify the snapshots which screened common reference compounds both by e-Pharmacophore models as well as docking. The same number (that a pharmacophore model screened) of top scoring compounds from the corresponding docking results were taken out and compared with the compounds screened by the respective e-Pharmacophore models (Figure 4), and the compounds screened by both the methods were identified. It was observed that in case of most of the MD snapshots, the pharmacophore model generated from the snapshot and docking screen a large number of common reference ligands as top scoring hits. So, this consistency was used as a criterion for selection of the models to be considered for virtual screening.

Five models generated from the snapshots of the MD trajectories of the model systems **E-SAHC-D** at 15 and 20 ns, **E-SAM** at 30 ns, and **E-SAHC** at 10 and 35 ns and were selected as the best models as the number of common compounds screened by the pharmacophore model and docking was more than 10. Table 1 lists the details of these five best models, and Figure 5 shows the docking scores and pharmacophore fitness scores of the reference compounds with these snapshots/models. The complete details of the pharmacophore matching and all components of docking scores for all the commonly matched compounds are given in Tables S5−S9 and Tables S10− S14, respectively. These models were also found to screen very small number of

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

noninhibitors with lower fitness scores as compared to the reference inhibitors. This shows the selectivity of the models to screen the true inhibitors of CmaA1. Table 2 shows a comparison of abilities of the selected models to screen the inhibitors and noninhibitors and a comparison of their fitness scores. The ability of the models to screen a majority of the inhibitors and a small fraction of noninhibitors further validate the models proposed here.

The models obtained from the snapshots **E-SAHC-D** at 15 and 20 ns have similar features and screen the same set of compounds. The reference compounds C1, C3, and C12 have highest fitness scores for the model generated from the snapshot **E-SAHC-D** at 15 ns and the number of common top scored compounds by pharmacophore screening and docking for this snapshot is 10. Similarly, the model generated from the snapshot **E-SAHC-D** at 20 ns matches the compounds C7, C9, and C11 with the highest fitness score and the number of common compounds was 13. The model generated from the snapshot **E-SAM** at 30 ns screens the compounds C2, C4, and C13 with highest score and the snapshot produced the highest docking score for C23. The number of common top scored compounds by pharmacophore screening and docking for this snapshot is 11. The model generated from the snapshot **E-SAHC** at 10 ns has 12 common compounds in the top scored hits of pharmacophore screening and docking. The reference compounds C6 and C17 have their highest fitness scores with this model. The pharmacophore model from the snapshot **E-SAHC** at 35 ns has 13 common compounds as the top scored matches for both pharmacophore screening and docking. Figure 6 shows these five best models along with the active site residues associated with the pharmacophoric features. Figure 7 depicts SAHC/SAM mapped to the five pharmacophore models and also the interactions of SAM/SAHC with the respective active site residues that give rise to the features.

Figure 8 shows the most active compound C1 mapped to the selected e-Pharmacophore models and also the interactions made by the compound with the active site residues of CmaA1 obtained from docking. The reference compounds were found to interact with the associated active site residues the same way as they are matched to the pharmacophore models. We propose the e-Pharmacophore models obtained from the snapshots **E-SAHC-D** at 15 and 20 ns, **E-SAM** at 30 ns, and **E-SAHC** at 10 and 35 ns as the best models that can be used for screening compound databases. Further studies on large scale virtual screening based on the models obtained here are in progress.

CONCLUSIONS

In this study structure based e-Pharmacophore models generated from the snapshots obtained from the MD simulation trajectories of five model systems of mycobacterial CmaA1 representing various stages of cyclopropanation and the reported crystal structure of CmaA1. The performance of these pharmacophore models were validated by mapping 23 thiacetazone analogues showing CmaA1 inhibitory activities (MIC) in the 0.0125 to 12.5 μg/mL range. The e-Pharmacophore models generated from the MD snapshots were able to screen up to 17 reference compounds. The models generated by considering the flexible nature of the protein were able to screen the reference compounds more efficiently as compared to the models generated from the static crystal structure indicating the need of incorporating receptor flexibility in virtual screening. We found that at least one model generated from the MD snapshots is able to screen up to 16 out of 23 reference compounds supporting the pre-existing receptor conformational ensembles theory. The models were also used to screen a set of 1398 noninhibitors of *M. Tb.* in order to opt out the possibilities of the selected models screening some false positives. The models were further validated by comparing the hits screened by the e-Pharmacophore model and docking the compounds with the respective snapshots. The five best models have been proposed based on the number of common compounds screened by both pharmacophore screening and docking methods for virtual screening. Further studies to compare the present results with ligand-based pharmacophore models, and virtual screening approaches are expected to be fruitful in identifying potential lead compounds for CmaA1 inhibition.

ASSOCIATED CONTENT

Supporting Information

Tables S1−S7 giving the details (number of compounds matched, docking, and fitness scores, etc.) of all the models. This material is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Authors

*E-mail: deva@iiit.ac.in (U.D.P.).

*E-mail: gnsastry@gmail.com (G.N.S.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.C. thanks Department of Science and Technology (DST), New Delhi, for financial assistance through an INSPIRE fellowship. U.D.P. thanks Department of Biotechnology (DBT) for financial assistance. G.N.S. thanks CSIR, New Delhi, for financial support in the form of XII five year project (GENESIS).

REFERENCES

(1) Cordeiro, D. C. J.; Malheiro, D.; Cadinha, S.; Duarte, R. Adverse Drug Reactions to All First-Line Antituberculosis Regimen and Ethionamide: A Successful Case Report. *Chest* **2014**, *145* (3 Suppl.), 94A−94A.

(2) Arcus, V. L.; Lott, J. S.; Johnston, J. M.; Baker, E. N. The Potential Impact of Structural Genomics on Tuberculosis Drug Discovery. *Drug Discovery Today* **2006**, *11*, 28−34.

(3) Lamichhane, G. Novel Targets in M. Tuberculosis: Search for New Drugs. *Trends Mol. Med.* **2011**, *17*, 25−33.

(4) Anand, P.; Sankaran, S.; Mukherjee, S.; Yeturu, K.; Laskowski, R.; Bhardwaj, A.; Bhagavat, R.; Brahmachari, S. K.; Chandra, N. OSDD Consortium. Structural Annotation of Mycobacterium Tuberculosis Proteome. *PloS One* **2011**, *6*, No. e27044.

(5) Szczepan, J.; Andrzej, S.; Katarzyna, K. How Mycobacterium Tuberculosis Subverts Host Immune Responses. *BioEssay.* **2008**, *30*, 943−954.

(6) Badrinarayan, P.; Sastry, G. N. Virtual Screening Filters for the Design of Type II P38 MAP Kinase Inhibitors: A Fragment Based Library Generation Approach. *J. Mol. Graph. Modell.* **2012**, *34*, 89−100.

(7) Badrinarayan, P.; Sastry, G. N. Virtual High-Throughput Screening in New Lead Identification. *Comb. Chem. High Thr. Scr.* **2011**, *14*, 840−860.

(8) Badrinarayan, P.; Sastry, G. N. Sequence, Analysis of P38 MAP Kinase: Exploiting DFG-Out Conformation as a Strategy to Design New Type II Leads. *J. Chem. Inf. Modell.* **2011**, *51*, 115−129.

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

(9) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. Virtual Screening in Drug Discovery − A Computational Perspective. *Curr. Protein Pept. Sci.* **2007**, *8*, 329−351.
(10) Badrinarayan, P.; Sastry, G. N. Rational Approaches towards Lead Optimization of Kinase Inhibitors: The Issue of Specificity. *Curr. Pharm. Des.* **2013**, *19*, 4714−4738.
(11) Villar, H. O.; Yan, J.; Hansen, M. R. Using NMR for Ligand Discovery and Optimization. *Curr. Opin. Chem. Biol.* **2004**, *8*, 387−391.
(12) Blundell, T. L.; Patel, S. High-Throughput X-Ray Crystallography for Drug Discovery. *Curr. Opin. Pharmacol.* **2004**, *4*, 490−496.
(13) Amzel, L. M. Structure-Based Drug Design. *Curr. Opin. Biotechnol.* **1998**, *9*, 366−369.
(14) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.
(15) Jiang, F.; Kim, S. H. Soft Docking": Matching of Molecular Surface Cubes. *J. Mol. Biol.* **1991**, *219*, 79−102.
(16) Walls, P. H.; Sternberg, M. J. E. J. New Algorithm to Model Protein-Protein Recognition based on Surface Complementarity: Applications to Antibody-Antigen Docking. *Mol. Biol.* **1992**, *228*, 277−297.
(17) Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A. Protein Folding in the Landscape Perspective: Chevron Plots and Non-Arrhenius Kinetics. *Proteins: Struct.,Funct., Bioinf.* **1998**, *33*, 74.
(18) Gschwend, D. A.; Good, A. C.; Kuntz, I. D. Molecular Docking towards Drug Discovery. *J. Mol. Recognit.* **1996**, *9*, 175−186.
(19) Leach, A. R. J. Ligand Docking to Proteins with Discrete Side-Chain Flexibility. *Mol. Biol.* **1994**, *235*, 345−356.
(20) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. J. The Sensitivity of the Results of Molecular Docking to Induced Fit Effects: Application to Thrombin, Thermolysin and Neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547−562.
(21) Bouzida, D.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Rose, P. W.; Verkhivker, G. M. Computer Simulations of Ligand−Protein Binding with Ensembles of Protein Conformations: A Monte Carlo Study of HIV-1 Protease Binding Energy Landscapes. *Int. J. Quantum Chem.* **1999**, *72*, 73−84.
(22) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.
(23) Cavasotto, C. N.; Abagyan, R. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209−225.
(24) Schapira, M.; Abagyan, R.; Totrov, M. Nuclear Hormone Receptor Targeted Virtual Screening. *J. Med. Chem.* **2003**, *46*, 3045−3059.
(25) Daeyaert, F.; de Jonge, M.; Heeres, J.; Koymans, L.; Lewi, P.; Vinkers, M. H.; Janssen, P. A. A Pharmacophore Docking Algorithm and Its Application to The Cross-Docking of 18 HIV-NNRTI's in Their Binding Pockets. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 526−533.
(26) Case, D. A. Normal Mode Analysis of Protein Dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285−290.
(27) Hayward, S.; Go, N. Collective Variable Description of Native Protein Dynamics. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223−250.
(28) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.
(29) Totrov, M.; Abagyan, R. Flexible Ligand Docking to Multiple Receptor Conformations: A Practical Alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178−184.
(30) Carlson, H. A. Protein Flexibility is an Important Component of Structure-Based Drug Discovery. *Curr. Pharm. Des.* **2002**, *8*, 1571−1578.
(31) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a Dynamic Pharmacophore Model for HIV-1 Integrase. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(32) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *J. Phys. Chem. A* **1999**, *103*, 10213−10219.
(33) Moitessier, N.; Henry, C.; Maigret, B.; Chapleur, Y. J. Combining Pharmacophore Search, Automated Docking, and Molecular Dynamics Simulations as A Novel Strategy for Flexible Docking. Proof of Concept: Docking of Arginine-Glycine-Aspartic Acid-Like Compounds into the Avβ3 Binding Site. *Med.Chem.* **2004**, *47*, 4178−4187.
(34) Meagher, K. L.; Carlson, H. A. Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.
(35) Meagher, K. L.; Carlson, H. A. Solvation Influences Flap Collapse in HIV-1 Protease. *Proteins: Struct.,Funct., Bioinf.* **2005**, *58*, 119−125.
(36) Damm, K. L.; Carlson, H. A. Exploring Experimental Sources of Multiple Protein Conformations in Structure-Based Drug Design. *J. Am. Chem. Soc.* **2007**, *129*, 8225−8235.
(37) Lerner, M. G.; Bowman, A. L.; Carlson, H. A. Incorporating Dynamics in E. Coli Dihydrofolate Reductase Enhances Structure-Based Drug Discovery. *J.Chem. Info. Modell.* **2007**, *47*, 2358−2365.
(38) Meagher, K. L.; Lerner, M. G.; Carlson, H. A. Refining the Multiple Protein Structure Pharmacophore Method: Consistency across Three Independent HIV-1 Protease Models. *J. Med. Chem.* **2006**, *49*, 3478−3484.
(39) Lexa, K. W.; Carlson, H. A. Protein Flexibility in Docking and Surface Mapping. *Q. Rev. Biophys.* **2012**, *45*, 301−343.
(40) Anuradha, A.; Trivelli, X.; Guérardel, Y.; Dover, L. G.; Besra, G. S.; Sacchettini, J. C.; Reynolds, R. C.; Coxon, G. D.; Kremer, L. Thiacetazone, An Antitubercular Drug that Inhibits Cyclopropanation of Cell Wall Mycolic Acids in Mycobacteria. *PLoS One.* **2007**, *12*, e1343.
(41) Choudhury, C.; Priyakumar, U. D.; Sastry, G. N. Molecular Dynamics Investigation of the Active Site Dynamics of Mycobacterial Cyclopropane Synthase during Various Stages of the Cyclopropanation Process. *J. Struct. Biol.* **2014**, *187*, 38−48.
(42) Huang, C.; Smith, C. V.; Glickman, M. S.; Jacobs, W. R., Jr.; Sacchettini, J. C. Crystal Structures of Mycolic Acid Cyclopropane Synthases from Mycobacterium tuberculosis. *J. Biol. Chem.* **2002**, *277*, 11559−11569.
(43) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; WuX, M.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Timulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1615.
(44) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(45) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−316.
(46) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.
(47) Salam, N. K.; Nuti, R.; Sherman, W. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *J. Chem. Inf. Model.* **2009**, *49*, 2356−2368.
(48) *LigPrep*, version 2.5; Schrödinger, LLC, New York, NY, 2012.
(49) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860

Screening. 1. Methodology and Preliminary Results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647−671.
(50) Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. Multiple Diverse Ligands Binding at a Single Protein Site: A Matter of Pre-Existing Populations. *Protein Sci.* **2002**, *11*, 184−197.
(51) Van Regenmortel, M. H. Molecular Recognition in The Post-Reductionist Era. *J. Mol. Recognit.* **1999**, *12*, 1−2.
(52) Salam, N. K.; Nuti, R.; Sherman, W. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *J. Chem. Inf. Modell.* **2009**, *49*, 2356−2368.
(53) Bosshard, H. R. Molecular Recognition by Induced Fit: How Fit is the Concept? *Physiology* **2001**, *16*, 171−173.
(54) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−796.

DOI: 10.1021/ci500737b
J. Chem. Inf. Model. 2015, 55, 848−860