Hybrid Pharmacophore Approach as a Novel Strategy to identify Anti-*Mycobacterial* Chemotypes: Investigation with DapB Enzyme

Chinmayee Choudhury[1],

Anshu Bhardwaj[2*]

[1]Department of Biochemistry, All India Institute of Medical Sciences, Basni, Jodhpur, Rajasthan, India

[2]Bioinformatics Centre, CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India

*Corresponding author

Email: anshu@imtech.res.in

Abstract

The current study examines the efficacy of dynamics-based hybrid pharmacophore models (DHPM) based on newly explored interaction features, as tools to screen potential inhibitors of *Mycobacterium tuberculosis* (*Mtb*)-DapB, a validated target essential for L-Lysin biosynthesis. Molecular dynamics (MD) simulations were performed on *Mtb*-DapB models, generated from a reported crystal structure by linking the cofactor (NADH) and substrate mimetic inhibitor (2,6-PDC), thereby creating a hybrid molecule (*HM*). Comparative investigation of the dynamics of interactions made by NADH, 2,6-PDC and *HM* from the MD trajectories revealed a number of stable interactions between *HM* and the hinge region residues leading to significant alterations of the *Mtb*-DapB structure. These newly identified interactions, translated into DHPM may be utilized as *Mtb*-DapB specific screening filters and provide new insights for structure activity relationships. To validate the applicability of the DHPM, about 10 million compounds compiled from the publicly available chemical space were screened using the DHPM as well as conventional pharmacophore models (based on the NADH/2,6-PDC). Systematic analyses of structures, physicochemical/ADMET properties and molecular docking of the 387 and 982 compounds screened by the DHPM and conventional models respectively demonstrate the capabilities of DHPM to screen structurally diverse chemical entities with better druglike properties and higher target binding potentials. The study demonstrates application of DHPM to predict 8 highly active anti-*Mtb* (MIC<=2μM) compounds whose mechanisms of action were previously unknown.

Key words

Pharmacophore; Virtual Screening; Molecular Dynamics; Docking; Mycobacterium tuberculosis

Introduction

Tuberculosis (TB) is the leading cause of death worldwide due to a single infectious agent according to the latest world health organization (WHO) reports [1,2]. The standard therapies of treating TB with a combination of several lines of antibiotics over a period of six to nine months [3, 4] have become ineffective due to emergence of drug resistant *Mycobacterium tuberculosis* (*Mtb*) strains [5]. The "End Tuberculosis Strategy" [1] of WHO calls for intensified research and innovation in TB drug discovery using multidisciplinary approaches to identify novel drug targets as well as fast and accurate techniques to design new chemical entities with higher potency to address the scourge of *Mtb*. Computational methods have become essential in last few decades not only to understand the drug-target interactions, but also for *in silico* screening of huge chemical libraries providing a fast and less expensive alternative to the traditional high throughput screening [6-11]. Recent literature has reported the identification of novel *Mtb* drug targets like DprE1, MmpL3, ATP, mycolic acid and Diaminopimelate (DAP) biosynthesis pathway enzymes etc that may be inhibited to tackle MDR and XDR strains [4, 12]. The DapB enzyme of the DAP biosynthetic pathway from H37Rv strain of *Mtb* (*Mtb*-DapB) is one of the validated drug targets [13, 14] as inhibition of this enzyme blocks the production of meso-diaminopimelate thus leading to inhibition of *de novo* lysine biosynthesis and peptidoglycan assembly, both of which are crucial for the survival of the pathogen [15] (**Fig 1a**).

Several groups made efforts to identify inhibitors of enzyme *Mtb*-DapB by exploring the potential of product analogues [16] as potential inhibitors, but have met with very limited success. 2, 6-PDC (Pyridine- 2, 6- dicaboxylate) and few other

heterocyclic aromatic product analogues have been identified with IC_{50} as high as 26 µM for *Mtb*-DapB [16]. Also, a number of sulphonamide inhibitors of *Mtb*-DapB which inhibited *Mtb*-DapB competitively with respect to the substrate 2,3-dihydrodipicolinic acid were identified with K_i values ranging from 7 to 48 µM [17]. This indicates that only substrate or product mimicry is not sufficient and new lead generation strategies should be employed exploring the key features of the binding sites of the enzymes. Molecular dynamics (MD) based pharmacophore models have emerged as quite powerful tools which not only account for the flexibility of the targets, but also help to identify novel key chemical features in the binding sites which is otherwise unexplored in the crystal structures and may be harnessed to design new inhibitors [18, 6, 7].

In this study, we have investigated dynamics-based hybrid pharmacophore modeling as a novel strategy to identify new anti-TB chemical space. MD simulations were performed on two model systems of *Mtb*-DapB to attain a detailed picture of the non-covalent interactions in the binding site responsible for stable cofactor and substrate binding and explore new regions of the binding site that may be targeted for effective enzyme inhibition. MD based e-pharmacophore models were generated representing the stable interactions of 2,6-PDC (a mimetic of the natural substrate of *Mtb*-DapB) and NADH (natural cofactor of *Mtb*-DapB) individually with their respective binding sites as well as hybrid e-pharmacophore models which represented interactions of both 2,6 PDC and NADH. A comprehensive dataset of above 10 million compounds compiled from publicly available chemical space was screened using these models. The structural and physicochemical properties of compounds screened by the hybrid models and the substrate/cofactor-based models were compared to examine the abilities of the hybrid models to identify a newer chemical

space with better druglike properties and higher binding affinities with the target. Although in this article, we have applied the dynamics-based hybrid pharmacophore strategy for *Mtb-DapB*, it can also be tested with other drug targets.

Material and Methods

Model Systems and Molecular Dynamics (MD) Simulations

Two different model systems of *Mtb*-DapB were considered for our study. The initial coordinates for the first model system **(DapB-N-P)** were taken from the crystal structure 1P9L [19], binding NADH and 2,6-PDC. The second model system **(DapB-Hyb)** was generated by connecting the nicotinamide part of NADH with 2,6-PDC of 1P9L using a propyl linker (**Fig 1b**) thereby creating a *hybrid molecule (HM).* The two systems were subjected to MD simulations using AMBER12 package [20]. The protein parts were prepared implementing AMBER ff99SB [21] force field while the ligands were prepared using General AMBER Force Field (GAFF) [22]. The partial atomic charges for NADH and PDC were derived using the antechamber module [23] by semi-empirical AM1 method with bond charge correction (BCC) model. A 1000-step steepest descent (SD) minimization was carried out on both the system to get rid of nonphysical contacts. Both the systems were then solvated in a cubical TIP3P water box and were neutralized by adding appropriate number of counter ions. A gradual heating from 0 to 300K was performed for 50 ps followed by 500 steps of SD and 500 steps of conjugate gradient energy minimizations. The systems were subjected to constant pressure equilibration (300K and 1 atm pressure) of 40 ns. Finally, the production run of 40 ns was performed under NPT ensemble 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 2 ps for analyses.

Generation of e-pharmacophore models

Eight snapshots were extracted at every 5 ns interval from each of the MD trajectories of the two model systems of *Mtb*-DapB, which were further used 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 ligands (NADH and 2,6-PDC in case of **DapB-N-P** and *HM* in case of **DapB-Hyb**) and their interactions with the receptor (*Mtb*-DapB) were evaluated by using "Score in place" mode of the Glide module of Schro□dinger molecular modeling suite [24, 25]. Default settings, with the option to output Glide extra precision (XP) [25] descriptor information, were used for the scoring. The resulting protein−ligand complexes along with the XP energy terms for hydrophobic enclosure, hydrophobically packed correlated hydrogen bonds, electrostatic rewards, π-π stacking, cation-π, and other interactions were then submitted to the Phase [26] module of Schrodinger to generate energy-based pharmacophore (e-Pharmacophore) models [27]. Each interaction is represented by a pharmacophore feature site and is assigned an energetic value equal to the sum of the Glide XP contributions of the atoms comprising the site. Then, the sites are ranked based on the energetic terms. Excluded volumes are added representing the receptor atoms surrounding the ligands. E-pharmacophore models generated using interactions of DapB with NADH and PDC (from the structures extracted from the MD trajectory of **DapB-N-P**) and the *HM* (from the structures extracted from the MD trajectory of **DapB-Hyb**) were named as N-type, P-type and H-type models respectively. Interactions made by only the portion

of *HM* excluding the ADP part (*HM**, **Fig 1**) were used for H type e-pharmacophore model generation.

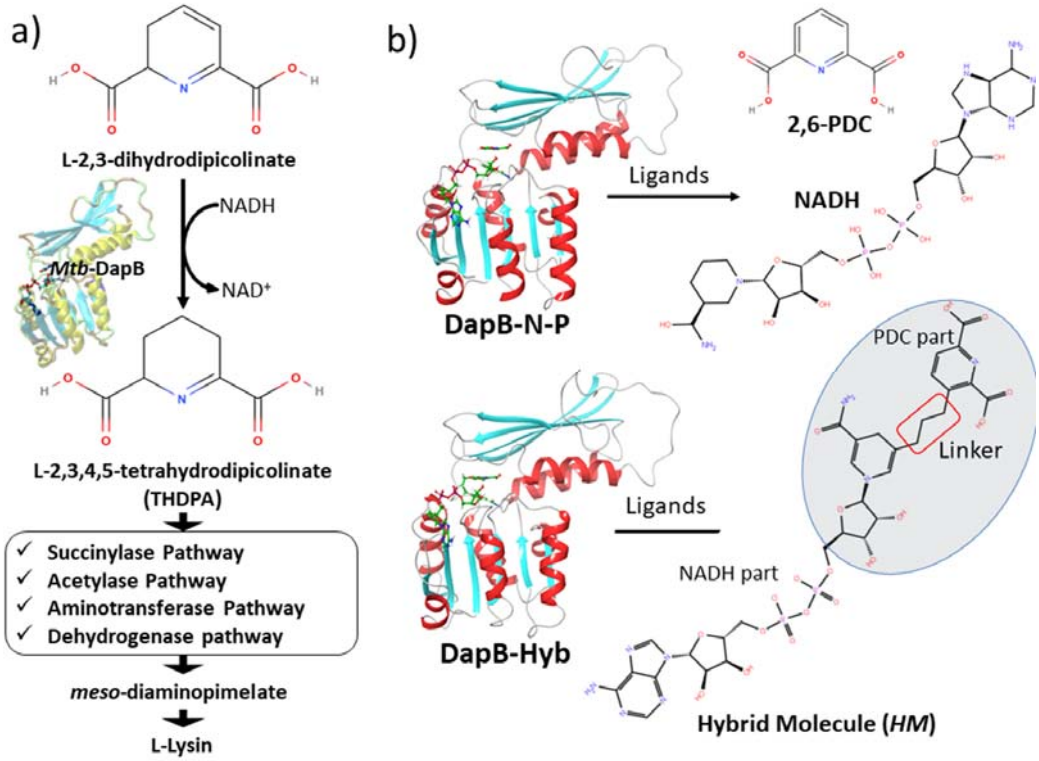

Fig 1. Function and model systems of Mtb-DapB. a. Catalytic function of *Mtb*-DapB in the Lysine synthesis pathway, **b.** Model systems considered for our study, **DapB-N-P:** Crystal structure of *Mtb*-DapB (1P9L), binding NADH and 2,6-PDC and **DapB-Hyb:** A model generated from 1P9L binding a *hybrid molecule(HM)*, created by connecting NADH and 2,6-PDC by a simple linker. The shaded portion of the hybrid molecule (*HM**) has been used for hybrid pharmacophore model generation.

Advanced pharmacophore screening

A large dataset comprising of 6886 molecules from DrugBank (1516 approved, 4788 experimental, 504 investigational and 78 nutraceutical drugs), 19500 ligands reported in PDB, 177000 molecules with activity<10μM reported in ChEMBL and 10639400 molecules from Zinc Drug-Like subset was considered for our screening study. This huge chemical space was preliminarily screened by applying the first level ligand based ElectroShape filters [28] implemented in the SwissSimilarity search server [29] taking NADH, 2,6-PDC and *HM** as query molecules. A unique set

of 6781 molecules screened by ElectroShape method using all the three query molecules were further screened by the dynamics structure-based e-Pharmacophore models. All these 6781 compounds were minimized using the default parameters of LigPrep [30] module of the Schrodinger Suite and five lowest energy conformers were retained for each compound. Screening was performed using the "Advanced Pharmacophore Screening" option of Phase module of Schrodinger Suite [26]. Five conformations per rotatable bond were generated for each ligand and maximum number of conformations per compound was assigned to be 100. A thorough 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 molecule must match was assigned to be 5 for the N-Type and H-type models, while it was 2 the P-type models. Among many conformers of a ligand, the one with the best fitness score (S) evaluated by a specific fitness function (S1 Equation) [27] was retained for each compound. The compounds screened by the N-, P- and H-type models were named as **N-set**, **P-set** and **H-set** respectively. Various cheminformatics analyses were performed to compare the structural, physicochemical and ADMET properties of the three sets of molecules.

Docking

The **N-set** and **H-set** molecules were docked to the MD snapshots obtained from DapB-N-P and DapB-Hyb respectively taking the same grids generated for e-pharmacophore generation (described in the '**Generation of e-pharmacophore models**' section). Flexible Glide XP docking calculations were performed to access the binding mode, interactions and affinities in terms of Glide XP score and ligand efficiencies of the **N-set** and **H-set** molecules. Evaluation of Glide score and ligand efficiency scores are described in S2 and S3 Equations [25]. Best 10 docking poses

were generated for each ligand and subjected to post docking minimization and then, 3 best energy poses per ligand were retained for analyses. **Fig 2** describes the overall methodology followed in this study.

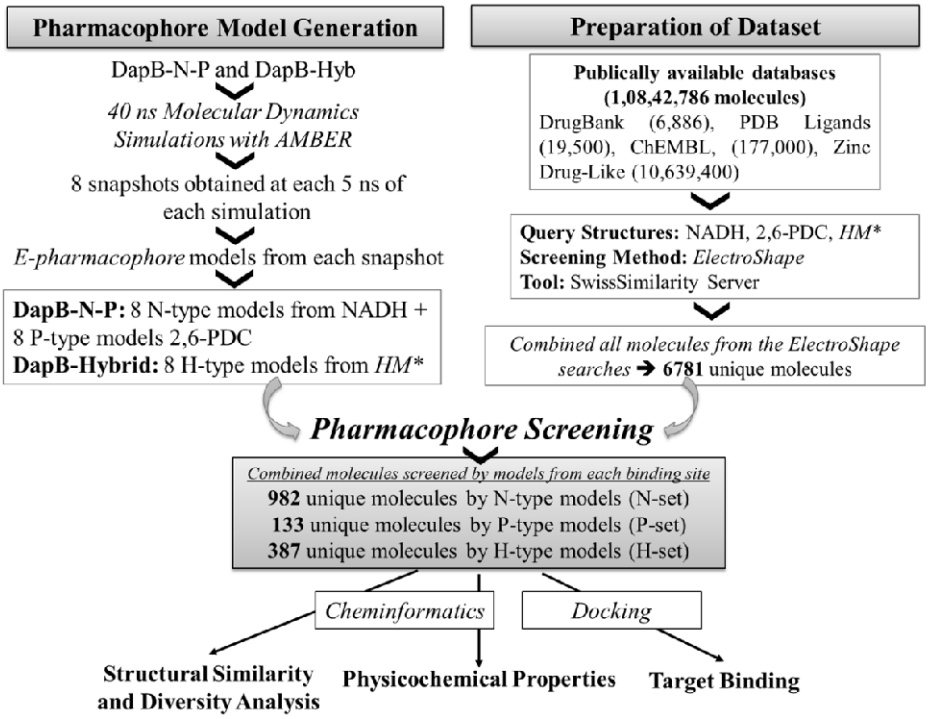

Fig 2. Systematic representation of the workflow followed in the study.

Screening of *Mtb* active (MA) compound library

A library of 2950 antitubercular compounds was carefully curated from literature [31] and ChEMBL database which is named as *Mtb* active (MA) compound library. These compounds have reported low micro molar MICs (Minimum inhibitory concentration) as low as 2 µM in the whole cell assays on *Mtb*. These compounds were prepared and screened against the eight H-type models with using the same parameters as described in the '**Advanced pharmacophore screening**' section. The compounds screened by each H-type model (compounds that matched at least 5 pharmacophore features) were docked to the respective DapB-Hyb snapshots (grids

generated for e-pharmacophore modelling were used) with the same Glide XP parameters as described in the **'Docking'** section.

Results and Discussion

Overall Structure and dynamics of *Mtb*-DapB in DapB-N-P and DapB-Hyb

The *Mtb*-DapB is a homo-tetramer of 245-residue monomers (**Fig 3a**). Two major domains and two short hinge regions connecting them comprise each monomer. The N-terminal domain (1-106 and 216-245) contains six β-strands (β1-β5 and β10) and four α-helices (α1-α3, α6). The C-terminal domain (107-211) consists of four βstrands (β6-β9) along with two α-helices (α4 and α5) giving rise to a mixed αβ-sandwich arrangement and a long loop (L8,156-179). Two short loops L4(103-106) and L10(212-215) connect two domains and act as hinge regions for the domain movements. The ADP part of NADH is embedded in a solvent exposed groove like region located in the N-terminal domain and extending to the hinge region, while the nicotinamide part is placed in the floor of a relatively less exposed cavity **C1** (**Fig 3b**) formed by residues from both N- and C-terminal domains. C-terminal side of C1 binds the substrate (dihydrodipicolinate) mimetic competitive inhibitor 2,6-PDC, which is stacked against the nicotinamide ring of NADH placed in the N-terminal side of C1. In this study we have considered two model systems; one of them is the crystal structure (1P9L) of *Mtb*-DapB reported in protein data bank and the other one was modelled by connecting the nicotinamide part of NADH with 2,6-PDC present in C1 using a propyl linker (*HM*). The stabilities of various interactions made by NADH, 2,6-PDC and *HM* with their respective binding sites have been thoroughly studied using the MD trajectories. The ways in which non-bonded interactions in the C1 differ in the two model systems were thoroughly examined.

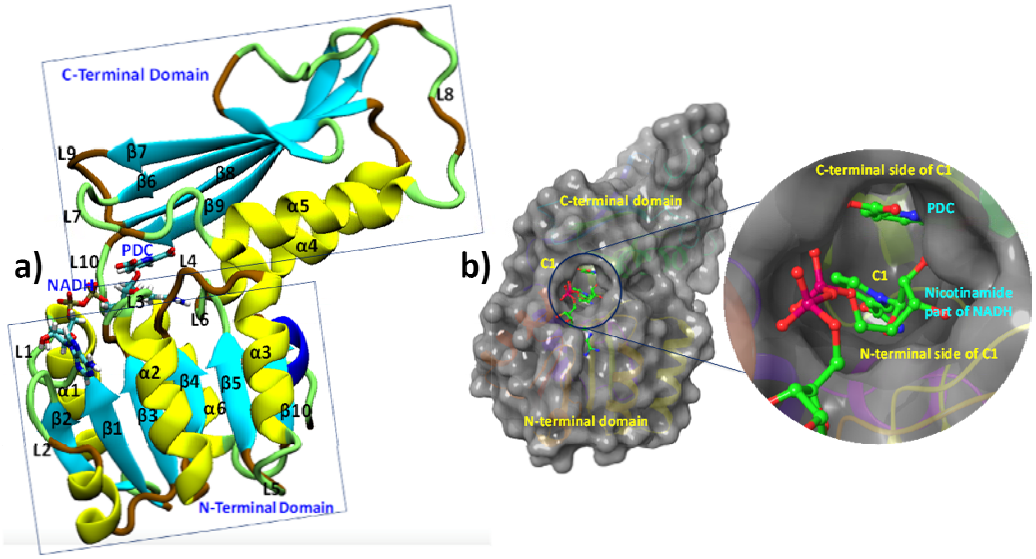

Fig 3. *Mtb*-DapB structure. a. Overall structural architecture of *Mtb*-DapB, **b.** NADH and substrate binding regions with a detailed picture of the C1 cavity.

From the MD simulation trajectories, we tried to study the structural differences in the *Mtb*-DapB models, when the ligands are free to move (as in DapB-N-P) and when they are connected (as in DapB-Hyb). We gave special attention to the binding pocket C1, which serves as a common region for binding of both the cofactor and substrate and thus can be targeted to block the binding of both the ligands. The graphs (**Fig S1a**) of root mean squared deviations (RMSD) of the overall structures during last 10 ns of the simulation do not show a fluctuation above 2.5 Å. RMSD of the ligands i.e., NADH, PDC and the *HM* also show a very stable profile during the simulations (**Fig S1b**), which suggests that the individual systems do not show major structural deviations during the simulations and are well equilibrated. Stable energy profiles of the systems also support the above observation (**Fig S1c**). Hence, the last 10 ns of both the trajectories were used for further analyses. The plots showing the interaction energies between the ligands and receptor (**Fig S1d**) during the last 10 ns of simulations clearly indicate that the *HM* binds to *Mtb*-DapB with a higher potential (more negative interaction energy) as compared to NADH and PDC individually.

Although this is an obvious observation owing to the size of the *HM* compared to NADH or PDC, we were keen to find out if the *HM* makes any new interactions (especially in the C1 region) or imparts any structural change in the *Mtb*-DapB binding site as compared to the original crystal structure which might contribute to its favorable binding.

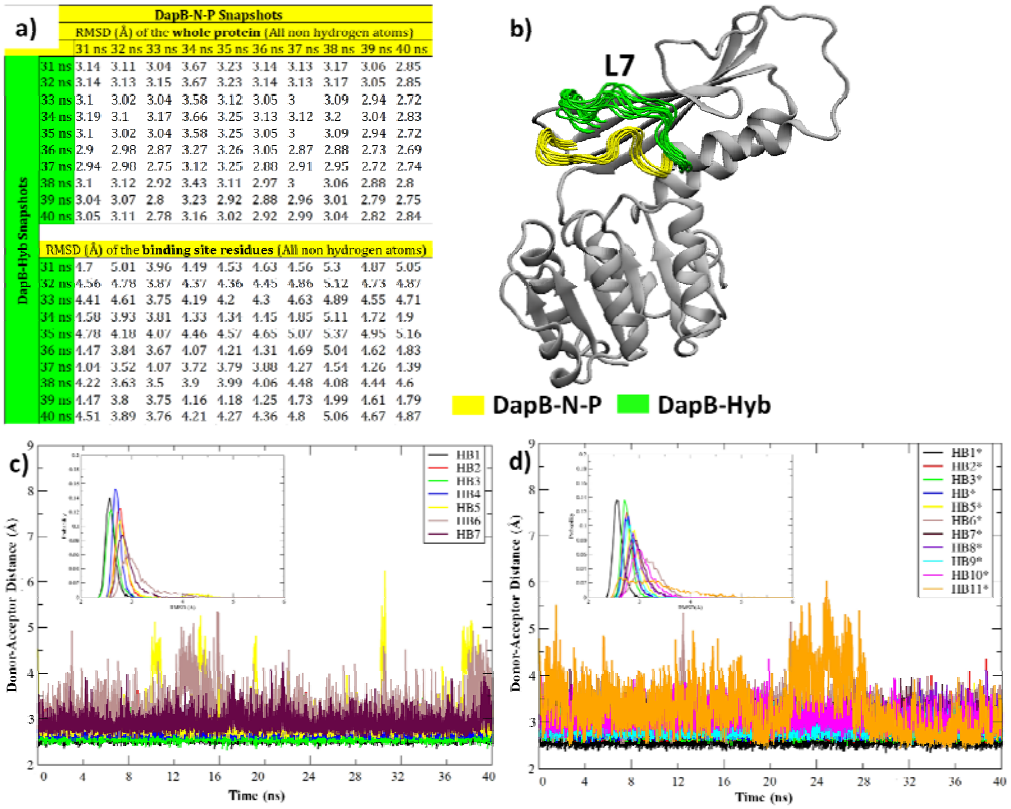

Fig 4. Comparative analyses of the structural features from the MD trajectories. a. Pair wise RMSD values computed by superposing each snapshot (coordinates saved at every 1 ns of the last 10ns of simulations) of the MD trajectories of **DapB-N-P** with those of **DapB-Hyb**. RMSD values for the overall structures as well as the binding site residues is shown. **b.** Elevation of the loop L7 of the C1 region in presence of *hybrid molecule* in the **DapB-Hyb** model, Distance profiles and probability distributions of HB Interactions formed by **c.** NADH and 2,6-PDC in DapB-N-P **d.** HB Interactions formed by *HM* in DapB-Hyb with >10% occupancy (highlighted in **Table S1)**. Other details of the HBs are given in Table S1.

To study these effects quantitatively, pair wise RMSD values were computed by superposing each snapshot (coordinates saved at every 1 ns of the last 10ns of simulations) of the MD trajectory of DapB-N-P with the corresponding snapshots of

DapB-Hyb. A similar matrix was also generated by superposing only the binding site residues of both the systems. These matrices (**Fig 4a**) indicated that there are significant structural differences between the two systems, especially in the binding site (C1) region. The L7 loop, which forms the C-terminal side of the binding site cavity C1 elevates about 3-4 Å in presence of the *HM* (**Fig 4b**) and this elevation mostly contributes towards the structural difference in the two systems. This observation proposes the possibility of distortion of *Mtb*-DapB structure upon binding of the *HM* which possesses combined interaction features of 2,6-PDC and nicotinamide part of NADH.

Analysis of receptor- ligand interactions in the binding site of *Mtb*-DapB from the MD trajectories

NADH and 2,6-PDC bind to *Mtb*-DapB mostly by making hydrogen bond (HB)s (**Fig S2**). In the crystal structure of *Mtb*-DapB (1P9L), the nicotinamide part of NADH makes interaction with G75, T77, A102, P103 and F105 of the C-terminal side of C1 while the ADP part makes interactions with G10, K11, V12, D33 and A34 of the N terminal domain. 2,6-PDC makes π-π stacking interactions with the nicotinamide ring of NADH and HB interactions with H133, K136, G142 and T143 of C-terminal region of C1, which are reported to be conserved among bacterial species [32] thereby affecting the specificity of the drugs that mimic dihydrodipicolinate. So, it is necessary to screen molecules which bind to C1 considering the interactions of both 2,6-PDC and the nicotinamide part of NADH and to capture such interactions, the stabilities of HBs made by NADH, 2,6-PDC and the *HM* were analyzed from MD trajectories of **DapB-N-P** and **DapB-Hyb** using the HB plugins of VMD [33].

During the 40 ns simulation of **DapB-N-P**, NADH and 2,6-PDC made 26 different HB interactions (**Table S1**) with various residues of *Mtb*-DapB out of which only 7

interactions showed more than 10% occupancy (highlighted in **Table S1**, **Fig 4c**). The most stable interactions made by the ADP part of NADH were with D33 and K11 while PDC made stable interactions with H132, H133 and K136 of C1 in **DapB-N-P**. It was found that the interactions formed by the nicotinamide part of NADH with A102, P103 and F105 are not stable (occupancy 0.06%, 0.14% and 0.60% respectively) throughout the simulation of **DapB-N-P**, although they are present in the crystal structure. In **DapB-Hyb**, out of 28 (**Table S1**, **Figs S2 and S3**), 11 interactions were found to be exhibiting more than 10% occupancy during the simulation (highlighted in **Table S1**, **Fig 4d**). The ADP part of the *HM* made stable interactions with K11, V12, G13 and T77, the nicotinamide part was found to be stably interacting with F105 of the hinge region. The PDC part of the *HM* made stable HBs with S141, H132, H133, K136, T143 of the C-terminal domain and R214 of the hinge region. Thus, in **DapB-Hyb** six new stable HBs, including two interactions with both the hinge region loops were formed, which were not found in **DapB-N-P**.

The C1 cavity of *Mtb*-DapB is formed by amino acid residues belonging to both C- and N-terminal domains as well as the hinge region loops connecting the two domains. When 2,6-PDC and NADH were not connected (in **DapB-N-P**), none of them made any interactions with the hinge region residues. 2,6 -PDC was bound to and moving mostly along with the C-terminal domain while NADH remained in the N-terminal domain during the simulations allowing a smooth inter-domain movement. But, when 2,6-PDC and the nicotinamide part of NADH (which occupies the N-terminal side of C1) were connected through a linker (as in **DapB-Hyb**), they were not able to move independently along with their respective domains as they formed many new interactions in the C1 including the hinge region residues (F105 and R214). Due to formation of these new interactions with the hinge region loops,

the free movements of the two domains was probably prevented resulting in significant structural alteration of the protein structure, which may act as a key point to inhibit the enzyme function. These observations prompted us to capture the new potential interactions of the part of the *HM* occupying the C1 cavity (*HM**) in the form of hybrid e-pharmacophore models (H-type) and test their potential to screen new chemical entities. Excluding the ADP part of *HM* in the H-type models would impart specificity by reducing the chance of screening molecules that resemble NADH or ADP, which may bind to a large number of other proteins in the host cells.

Generation and comparison of e-Pharmacophore models

The receptors are normally very flexible and exist in a number of conformations in their natural environment out of which some conformations are appropriate for binding the ligands. So, in this study we decided to consider multiple conformations of the binding pockets of *Mtb*-DapB given the possibility of diverse conformational states of the protein-ligand complexes leading to diverse interactions. Eight snapshots were collected from MD trajectories of each **DapB-N-P** and **DapB-Hyb** model systems and three types of e-pharmacophore models viz., N-, P- and H-type were generated based on the interaction of NADH, 2,6-PDC and the *HM** respectively. **Fig 5** shows one representative from each of the three types of pharmacophore models and **Tables 1, S2 and S3** and **Figs S4 and S5** give the details of all the models belonging to the three categories.

The e-pharmacophore approach uses the energy components of the receptor-ligand binding from Glide XP[25] to generate structure-based pharmacophore models. The models comprise six types of chemical features, viz., HB acceptor (A), HB donor

(D), hydrophobic sites (H), negative ionizable sites (N), positive ionizable sites (P) and aromatic rings (R).

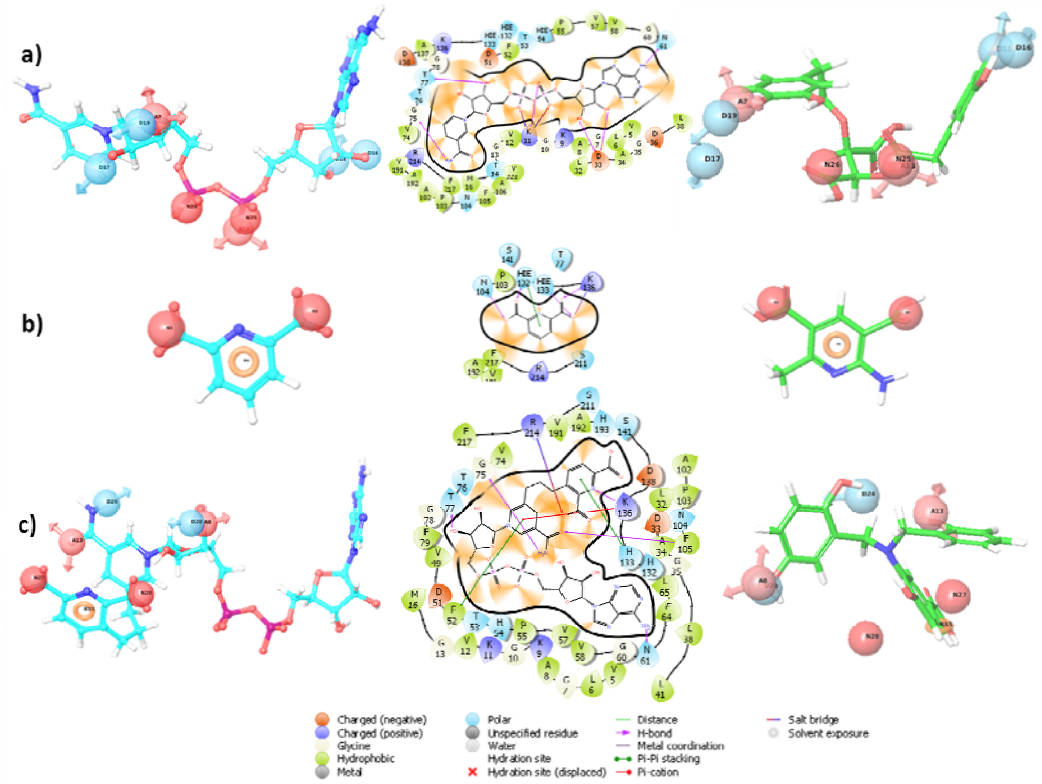

Fig 5. Representatives from each of the three types of pharmacophore models. a. N-type, **b.** P-type and **c.** H-type models along with the corresponding interactions from which they are originated and each model mapped to the respective screened ligand with highest fitness score. Colour codes for the pharmacophoric features are as follows. Cyan: D, Pink: A, Red: N, Blue: P, Green: H and Orange: R. Same colour code for the features is followed for all the other figures.

In our study, the pharmacophoric features were chosen considering the stabilities of the interactions as deciphered from the analysis of the MD trajectories as well as few instantaneous interactions formed in a particular snapshot as quantified by Glide XP. The N-type models possess 7-8 features. The D features near O_2B and O_3B (please refer **Fig S3** for atom nomenclatures) of ADP sugar are formed because of the H-bonding with the residues D33 and occasionally G7, while an additional D feature is present in some N-type models near N_6A of ADP part with N61.

Table 1. Details of one representative from each type i.e., N-, P- and H- type dynamics-based e-pharmacophore models.

Feature Label	Type	Score	X	Y	Z	Source
N-type model						
A7	A	-1.6	42.881	29.313	26.909	HB
D19	D	-1.23	42.379	30.142	26.996	HB
A13	A	-1	39.643	22.201	23.596	HB
N25	N	-1	40.165	22.3935	24.664	Ionic+HB
D16	D	-0.8	44.02	16.508	24.822	H
D17	D	-0.8	40.285	30.883	24.995	HB
D18	D	-0.72	43.871	17.81	23.024	HB
D23	D	-0.28	45.041	32.869	22.37	HB
P-type model						
N3	N	-1.25	42.37	27.368	23.611	HB
N2	N	-0.07	42.8005	33.0355	21.2585	Ionic+HB
R4	R	-0.73	42.0342	29.6038	20.9	Ring (Chemscore)
H-type model						
D20	D	-1.55	43.126	27.147	25.335	HB
N28	N	-1.25	39.671	28.1175	21.369	HB
N27	N	-1.25	40.006	20.2165	21.9075	Ionic+HB
D24	D	-1.03	44.185	15.296	21.783	HBond
R33	R	-0.86	39.9987	30.31	18.6402	Ring (Chemscore)
A8	A	-0.76	43.563	26.317	25.074	HB
A6	D	-0.66	44.35	14.466	23.826	HB

One or two N features were found near the O_1N and O_2A of the phosphate groups of ADP part of NADH. These atoms also make strong and stable HBs with K11 and give rise to a constant A feature in almost all N-type Models. As discussed in the previous section, the O_2D/O_3D and N_7N of the nicotinamide part of NADH make stable H-bonding with T77 and G75 and A102 respectively to give rise to the D and A features near C1 region in most of the N-type Models. 2, 6-PDC being a very small molecule, could give rise to the P Type e-pharmacophore models with only 3 to 4 features. One of them is a R feature representing the π-π interaction with nicotinamide ring of NADH. The strong electrostatic contribution of O_1, O_2, O_3 and O_4 towards the HB interactions with H133, K136, G142 and T143 give rise to two N features in all the P-type models and an A feature is present near the N_1 atom representing the HB

interaction with K136. However, The H-type models, derived from the MD trajectory of **DapB-Hyb** were generated by considering the inter molecular interactions of the C1 regions only. Hence, the most striking differences between the N- and H-type models are absence of features corresponding to the interactions of the ADP region in the later. Each H-type e-pharmacophore model comprises of seven features. The two N features near the carboxylic groups of the PDC part of the *HM** represent the salt bridge interactions with K136 and H132. The aromatic ring in PDC portion of the *HM* makes π-π stacking with H132 and the hinge region residue F217 and cation-π interactions with the hinge region residue R214 in few snapshots. This ring feature is present in H-type models and not in N-type models. HB interaction of -OH groups of the nicotinamide sugar with T77 gives rise to a D or an A feature in almost all models. The amide $-NH_2$ of the nicotinamide part of the hybrid region makes interactions with G75, A102 and occasionally D51 giving rise to the A and D features. Considering the striking differences in the number and spatial orientations of the pharmacophoric features in the three types of models, we screened a huge chemical space using these models and compared the structures, druglike properties and target binding affinities of molecules screened by each type of models.

Comparative analysis of the molecules screened by the three types of pharmacophore models

As the three types of pharmacophore models were significantly different, it was interesting to study the similarities and diversities of molecules screened by each of them (Set-N, Set-P and Set-H compounds as described in the method). As NADH is a very abundant metabolite in human, the N-type pharmacophore models generated from NADH may not screen compounds with high specificity for *Mtb*-DapB. At the same time 2,6-PDC being a very small molecule, gives P-type pharmacophore models

with only 3 to 4 features, which may fail to screen *Mtb*-DapB specific molecules. As observed from our MD simulation trajectories, interactions of the *HM*s with the hinge region residues bring about significant structural alterations in the binding site region, and none of the N- and P- types of model possesses features representing the hinge region interactions. So, the whole idea of generating the H-type models was to test them as tools to screen new compounds which not only bind to *Mtb*-DapB with high affinity and specificity, but also target the hinge region residues to alter the protein structure.

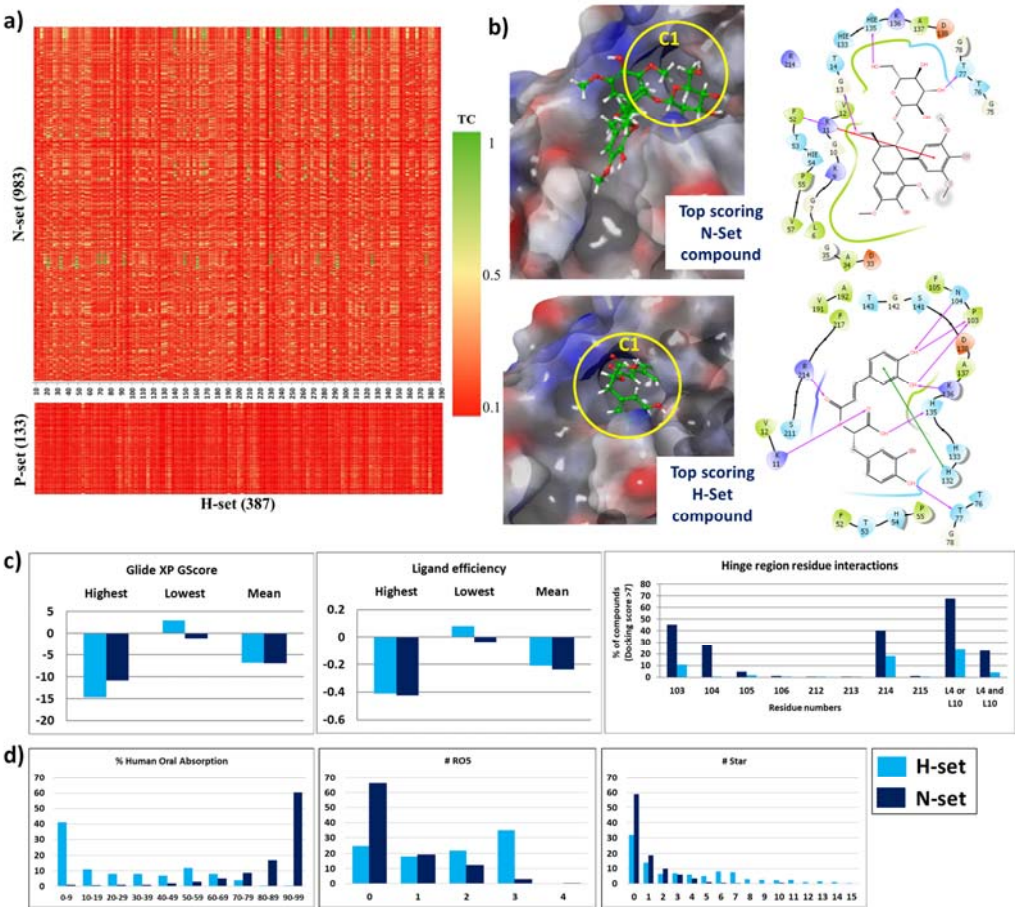

Fig 6. Cheminformatics analyses and molecular docking results. a. Similarity matrix of N- and P-Set compounds with the H-set compounds, **b.** Binding mode and interactions (please refer the legend given in **Fig 5**) of the top scoring N-set and H-set compounds, **c.** Docking score, ligand efficiency and percentage of N-set and H-set compounds those make interactions with the hinge region residues, **d.** Distribution of druglikelines indices #stars, #RO5 and human oral absorption for N-(Cyan) and H-set (Navy blue) compounds.

The 133 P-set compounds were mostly single ring and low molecular weight fragment like structures. So, they were excluded in the comparative analysis of properties considering their small size and lack of chemotype diversity. The N-type models screened 982 molecules while the H-type models screened 387 compounds.

As a preliminary comparison, the structural similarity between Set-N/Set-P compounds were evaluated as Tanimoto coefficient (TC) [34] using OpenBabel [35] and a small inhouse shell script. The results are presented as a color-coded matrix (**Fig 6a**), where a TC value from 0 (lowest) to 1 (highest) is coded by red through yellow to green. Thus, the red shades represent lower similarity and the yellow to green shades represent higher similarities. A quick look at the matrix shows the predominance of the red color indicating that the structures of the H set compounds are significantly different from those of the N- and P- set compounds. A near neighbor analysis was carried out to delineate the chemical diversity of the N- and H-set compounds with the ChemAxon program [36]. JKlustor was used for clustering and diversity analysis of chemical sets. The search returned 218 singletons for the 387 H-set compounds which is about 56.33%, while 340 singletons were obtained from the 982 N-set compounds counting only 34.5% of the total N-set compounds. These results indicate that the molecular diversity is higher in case of H-set compounds as compared to the N-set compounds.

In order to understand how the N-and H-set compounds bind to *Mtb*-DapB, XP docking calculations were performed, where the compounds from both the sets were docked to all the 16 snapshots obtained from the MD trajectories of DapB-N-P and DapB-Hyb. 168 out of the 387 (~43.5%) H-set compounds and 370 out of 982 (~37.4%) N-Set compounds were found to have Glide XP Gscore >7 with at least one snapshot of *Mtb*-DapB. Although, the range of scores for the N-set compounds seems

to be higher (-14.7 to 2.97) as compared to that of the H-set compounds (-10.85 to -1.19), the average score for the H-set compounds is slightly higher than the N-set compounds. As most of the scoring functions are biased towards larger ligands, we also considered the ligand efficiency indices [37] (**Figs 6c** and **S6**), which effectively normalize the docking scores by the heavy atom count and surface area eliminating the common bias of docking score in favor of larger ligands. The H-set compounds showed better average ligand efficiency than the N-set compounds. **Fig 6b** and the first two graphs of **Fig 6c** indicate that, both sets of compounds bind to *Mtb*-DapB with almost equal strength (H-set being little higher), but the mode of binding is different. It was observed that while most of the N-set compounds bound to the relatively more solvent exposed groove like region (where the ADP part of NADH binds in the crystal structure), the H-set compounds occupy the C1 cavity (**Fig 6b**). As identified from the analyses of MD trajectories of DapB-Hyb, interactions of *HM* with the two hinge region loops L4(103-106) and L10(212-215) were found to be crucial for structural alteration of *Mtb*-DapB. Hence, we also checked if the N- and H-set compounds interact with the hinge regions. Interestingly, we observed that about 68% of the H-set compounds formed interactions with at least one hinge region loop, while only 23% of N-set compounds show such interactions. 22% of H-set compounds were found to make interactions with both the hinge region loops while this number is as low as 4% in case of N-set compounds (**Fig 6c**, 3rd graph). This observation clearly indicates that the H-type pharmacophores are able to screen compounds that specifically interact with the hinge region residues in the C1 cavity and hence may impart better enzyme inhibitory effect.

Apart from the binding specificity, the druglikeliness of the N-and H-set compounds were also compared to further prioritize the H-type pharmacophore

models. **Fig S7** shows comparison of various physicochemical and drug like properties (**S1 List**) of the H- and N- sets of compounds calculated with the QuickProp module of Schrodinger Suite [38]. A quick look at the distribution of various structural properties of the N and H set compounds show that the N-set compounds are larger molecules having higher ranges of molecular weight, number of heavy atoms, molecular volume and solvent accessible surface area (SASA) as compared to the H set compounds. This is due to the higher inter feature distances in the N type pharmacophore models as compared to the H type models. However, when we compare the ring systems comprising the N and H set compounds, we find that the H set compounds have more number of rings, especially aromatic/hereroaromatic rings as all most all the H type models have a R feature, while the set N compounds possess slightly higher number of aliphatic rings. Thus, the π (carbon and attached hydrogen) component of the SASA (PISA) was found to be higher in case of the H set compounds as compared to the N set compounds. As discussed earlier, the N- and H-type models are derived from a more solvent accessible region and a relatively less accessible C1 region of the *Mtb*-DapB binding site respectively. Hence, interestingly the hydrophilic component of SASA (FISA) and Van der Waals surface area of polar nitrogen and oxygen atoms (PSA) of compounds screened by the respective models also follow a similar trend, that is, the N set compounds show higher values of PISA and PSA. About 40% and 45% of the N set compounds were found to have PSA and FISA values respectively which are beyond the recommended range[37] (S1 List) (7-200 for PSA and 7-330 for FISA, followed by 95% of known drugs) while these values are well within the recommended range in case of H set compounds. The number of HB donors and acceptors are below 5 and 10 for more than 95% of the H set compounds following the rule of five for druglikeliness, while a very small

fraction of N set compounds falls within these ranges. The solubility, bioavailability and the drugliklinesss scores were found to follow strikingly different trends in case of N and H set compounds. The #star descriptor indicates the number of property or descriptor values (such as molecular weight, dipole moment, ionization potential, electron affinity, SASA and its components, volume, HB donor and acceptor, globularity, solubility, lipophilicity, bioavailability, toxicity etc. (refer List S1 for detailed information)) that fall outside the 95% range of similar values for known drugs. A higher value range (about 40% molecules have a value >=5) of #stars (**Fig 6d**) for the N set compounds suggests they are less drug-like than H set molecules with few stars (only 1% molecules have a value >=5). Similarly, the #RO5 values (Number of violations of Lipinski's rule of five) are also higher for the N set compounds as compared to the H set compounds (**Fig 6d**) showing better druglikeliness of the later. The #metab descriptor is a predicted value representing number of likely metabolic reactions gives an estimation of the off-target interactions and toxicity of the compounds. The N set compounds show a higher number as compared to the H set compounds. Hence with all these comparative observations of the structural and physicochemical properties, and drug-likeliness scores, we can summarize that, the hybrid pharmacophore models lead to a structurally diverse and more druglike chemical space.

Application of the hybrid pharmacophore models to decipher the mechanism of action of anti-*Mtb* compounds

Considering the abilities of the hybrid pharmacophore models, they can have versatile applications. For example, they can be directly implemented to screen the huge chemical space to suggest potential binders of *Mtb*-DapB, which may act as start

points for inhibitors design, the spatial locations of the pharmacophoric features may also exploited for fragment based de novo drug design and scaffold hopping.

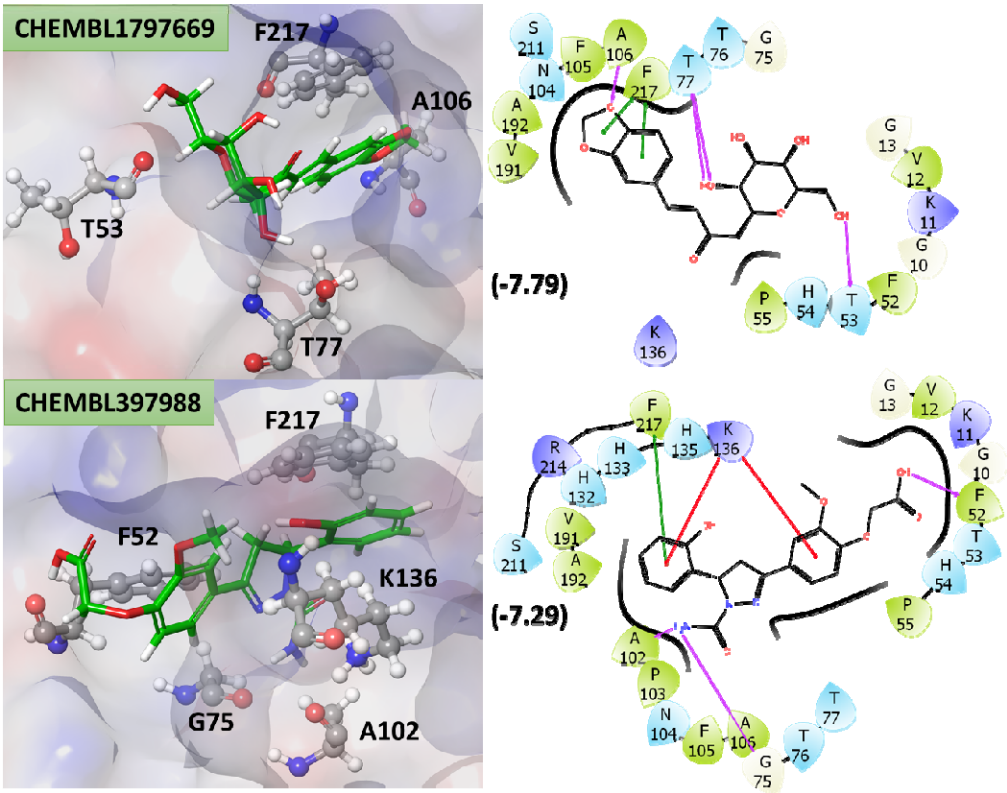

Fig 7. Interactions of two of the MA library compounds with the C1 cavity residues (please refer the legend given in **Fig 5**). The respective Glide XP docking scores are mentioned in the brackets.

In this study we demonstrate the application of the models to screen a library of known antitubercular compounds to find their possible mechanism of action. The MA library was created by extracting 2950 *Mtb* active compounds from ChEMBL and also manually from literature. These molecules have been reported to have ability to kill *Mtb* at concentration < 2 µM in the whole cell based essays, but their mechanism of action is not known. Screening this library with the eight H-type models returned a unique set of 31 compounds, which match five or more pharmacophoric features. These compounds were docked to the DapB-Hyb snapshots to estimate their binding potentials as well as analyze their interactions with the crucial residues identified from the MD studies. Eight compounds were found to have the XP docking score

above 7 (Table S4). These molecules make H-bonds, π-π, cation-π and salt bridge interactions with both the hinge region as well as other C1 residues (**Fig 7, Table S4**). Hence, we predict these molecules might be acting on *Mtb* cells through inhibition of *Mtb*-DapB by competitively binding to both the cofactor and substrate binding sites and/or restricting the inter-domain motions. However, these molecules need further experimental investigation to confirm their mechanism of action.

Conclusions

Most of the currently practiced rational drug design approaches are based on the interactions of the natural ligands such as the cofactors or known inhibitors with their respective binding sites. Many other interaction features of the binding site may also be explored when the structure of the target protein is known. The hybrid dynamic-pharmacophore model approach investigated in this manuscript is based on the observations that when a hybrid molecule binds to *Mtb*-DapB by making interactions with multiple binding sites of the enzyme, it results in significant alterations of its structure by restricting inter domain movements. The analyses of the MD trajectories of *Mtb*-DapB with the existing two co-crystalized ligands and a *HM* modelled by linking these two ligands indicated that, interactions of the *HM* with the hinge region residues bring about significant structural alterations in the binding site of *Mtb*-DapB. These newly identified interactions may be utilized as *Mtb*-DapB specific screening filters and provide new insight for structure activity relationships. We report these interactions in the form of hybrid e-pharmacophore models, which can be used as key filters in structure based *de novo* design/screening of *Mtb*-DapB inhibitors. The compounds screened by the hybrid models were found to specifically make interactions with the hinge region residues and have reasonably good binding affinities with *Mtb*-DapB as estimated from the docking studies. Also, the hybrid

pharmacophore models are capable of screening compounds which are highly diverse, structurally different and found to have better druglike properties than those screened by the models generated from NADH and 2,6- PDC individually. The study also demonstrates one of the many applications of these models by screening a library of compounds which are *Mtb* membrane permeable and show MIC values below 2μM to find if any of them acts through *Mtb*-DapB inhibition. Eight of them were found satisfying the pharmacophoric requirements of binding sites as well as making energetically favourable binding interactions. These findings may provide some clue for further experimental investigation to understand the mechanism of action of these compounds. Considering its broad applicability, we propose the hybrid e-pharmacophore modeling as a novel and effective strategy to venture into novel and potential anti-*Mtb*-DapB chemical space. Though in this article, we have discussed the dynamics-based hybrid pharmacophore strategy taking *Mtb*-DapB as an example, it can also be tested with a broad range of other drug targets.

Acknowledgements

CC thanks Department of Science and Technology, Govt. of India for financial support in the form of INSPIRE-Faculty award [DST/INSPIRE/04/2016/000732]. AB acknowledge financial support from OLP0753. CC and AB thankfully acknowledge Prof. Prasad V. Bharatham of National Institute of Pharmaceutical Education and Research, Mohali, for allowing to use Schrodinger software at his lab.

References

1. WHO Global Tuberculosis Report. http//www.who.int/tb/en/ (2017)
2. Balganesh, T. S., Alzari, P. M. & Cole ST. Rising standards for tuberculosis drug development. Trends Pharmacol. Sci. **29,** 576-581 (2008)

3. Gleeson, T. D. & Decker, C. F. Treatment of Tuberculosis. Disease-a-Month. **52(11-12),** 428-434 (2006).
4. Zumla, A. et al. Tuberculosis treatment and management-an update on treatment regimens, trials, new drugs, and adjunct therapies. Lancet Respir. Med. **3**, 220–234 (2015)
5. Prasad, R. & Kumar, D. Multi drug and extensively drug-resistant TB (M / XDR-TB) management: Current issues. CEGH Clin. Epidemiol. Glob. Heal.; **1**, 124–128(2013)
6. 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. **187**, 38-48 (2014)
7. 6. Choudhury, C., Priyakumar, U. D. & Sastry, G. N. Dynamics Based Pharmacophore Models for Screening Potential Inhibitors of Mycobacterial Cyclopropane Synthase. J. Chem. Inf. Model. **55**, 848-860 (2015)
8. Choudhury, C., Priyakumar, U. D. & Sastry, G. N. Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors J. Chem. Sci. **128**, 719-732 (2016).
9. Srivastava, H. K., Choudhury, C. & Sastry, G. N. The efficacy of conceptual DFT descriptors and docking scores on the QSAR models of HIV protease inhibitors. Med. Chem. **8.5**, 811-825 (2012)
10. Choudhury, C., Priyakumar, U. D. & Sastry, G. N. Structural and Functional Diversities of the Hexadecahydro□1H□cyclopenta [a] phenanthrene Framework, a Ubiquitous Scaffold in Steroidal Hormones. Mol Info. **35**, 145-157 (2016)

11. Gaur, A. S. et al. Assessing therapeutic potential of molecules: molecular property diagnostic suite for tuberculosis (MPDS^{TB}). J Chem Sci. **129**, 515-531. (2017)
12. Mdluli, K., Kaneko, T. & Upton, A. Tuberculosis drug discovery and emerging targets. Ann. NY Acad. Sci. **1323**, 56–75 (2014)
13. Cox, R. J. The DAP pathway to lysine as a target for antimicrobial agents. Nat. Prod. Rep. **13**, 29–43 (1996)
14. Cox, R. J., Sutherland, A. & Vederas, J. C. Bacterial diaminopimelate metabolism as a target for antibiotic design. Bioorg. Med. Chem. **8**, 843–871 (2000)
15. Usha, V., Lloyd, A. J., Lovering, A. L. & Besra G. S. Structure and function of Mycobacterium tuberculosis meso-diaminopimelic acid (DAP) biosynthetic enzymes. FEMS Microbiol. Lett. **330**, 10–16 (2012)
16. Singh S. P., Konwar, B. K., Bezbaruah, R. L. & Bora, T. C. Molecular docking and in silico studies on analogues of 2-methylheptyl isonicotinate with DHDPS enzyme of Mycobacterium tuberculosis. Med. Chem. Res. **22**, 4755–4765 (2013)
17. Paiva, A. M., Vanderwall, D. E., Blanchard J. S., Kozarich, J. W., Williamson J. M. & Kelly, T. M. Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. Biochim. Biophys. Acta. **1545**, 67–77 (2001)
18. Hendrychova, T., Berka. K., Navratilova, V., Anzenbacher, P. & Otyepka, M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. Curr. Drug Metab. **13**, 177–189 (2012)

19. Cirilli, M., Zheng, R., Scapin, G. & Blanchard, J. S. The Three-Dimensional Structures of the Mycobacterium tuberculosis Dihydrodipicolinate Reductase-NADH-2,6-PDC and -NADPH-2,6-PDC Complexes: Structural and Mutagenic Analysis of Relaxed Nucleotide Specificity. Biochemistry. **42**, 10644-10650 (2003)

20. Case, D. A. et al. AMBER 12, University of California, San Francisco. (2012)

21. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Struct. Funct. Bioinf. **65,** 712–725 (2006)

22. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P.A & Case, D. A. Development and testing of a general amber force field. J. Comput. Chem. **25**, 1157–1174 (2004)

23. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. J. Mol. Graph. Modell. **25**, 247–260 (2006)

24. Maestro, version 11.4 (Academic), Schrödinger, LLC, New York, (2017)

25. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R. & Halgren, T. A. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. J. Med. Chem. **49**, 6177–6196 (2006)

26. 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 Screening. 1. Methodology and Preliminary Results. J. Comput. Aid. Mol. Des. **20**, 647-671 (2006)

27. Salam, N. K., Nuti, R. & Sherman, W. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. J. Chem. Inf. Model. **49**, 2356–2368 (2009)

28. Armstrong, M. S. et al. ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. J. Comput. Aid. Mol. Des. **24**, 789-801 (2010)

29. Zoete, V., Daina, A., Bovigny, C. & Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. J Chem. Inf. Model. **56**, 1399-1404 (2016)

30. LigPrep Schrödinger, LLC, New York, NY. (2015)

31. Stanley, S. et al. Identification of novel inhibitors of M. tuberculosis growth using whole cell based high-throughput screening. Am. Chem. Soc. Chem. Biol.; **7**,1377–1384 (2013)

32. Scapin, G., Reddy, S. G., Zheng, Blanchard, J. S. Biochemistry; **36**, 15081-15088 (1997)

33. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. J. Mol. Graphics. **14,** 33-38 (1996)

34. Tanimoto, T. An Elementary Mathematical theory of Classification and Prediction. International Business Machines Corporation; 1-10 (1958)

35. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T. & Hutchison, G. R. Open Babel: An open chemical toolbox. J. Cheminf. **3,** doi: 10.1186/1758-2946-3-33 (2011)

36. JChem 17.22.0, ChemAxon, http://www.chemaxon.com. (2017)

37. Hopkins, A. L., Keserü, G. M., Leeson, P. D., Rees, D. C., Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. Nat. Rev. Drug Discov. **13,** 105-121 (2014)

38. QikProp, Shrödinger, LLC, New York, NY (2015).