Hybrid Dynamic Pharmacophore Models as Effective Tools to Identify Novel Chemotypes for Anti-tb Inhibitor Design: a Case Study With Mtb-dapb

Chinmayee Choudhury
Post Graduate Institute of Medical Education and Research, Chandigarh, India https://orcid.org/0000-0002-0152-133X

Anshu Bhardwaj (anshu@imtech.res.in)
Institute of Microbial Technology CSIR https://orcid.org/0000-0001-5830-2616

Research article

Keywords: Pharmacophore, Docking, Mycobacterium, Antimicrobial resistance, Molecular dynamics, Chemotype, Cheminformatics

Posted Date: June 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37081/v1

Abstract

Antimicrobial resistance (AMR) is one of the most serious global public health threats as it compromises the successful treatment of deadly infectious diseases like tuberculosis. New therapeutics are constantly needed but it takes a long time and is expensive to explore new biochemical space. One way to address this issue is to repurpose the validated targets and identify novel chemotypes that can simultaneously bind to multiple binding pockets of these targets as a new lead generation strategy. This study reports such a strategy, dynamic hybrid pharmacophore model (DHPM), which represents the combined interaction features of different binding pockets contrary to the conventional approaches, where pharmacophore models are generated from single binding sites. We have considered *Mtb*-DapB, a validated mycobacterial drug target, as our model system to explore the effectiveness of DHPMs to screen novel unexplored compounds. *Mtb*-DapB has a cofactor binding site (CBS) and an adjacent substrate binding site (SBS). Four different model systems of *Mtb*-DapB were designed where, either NADPH/NADH occupies CBS in presence/absence of an inhibitor 2, 6-PDC in the adjacent SBS. Two more model systems were designed, where 2, 6-PDC was linked to NADPH and NADH to form hybrid molecules. The six model systems were subjected to 200ns molecular dynamics simulations and trajectories were analysed to identify stable ligand-receptor interaction features. Based on these interactions, Conventional pharmacophore models (CPM) were generated from the individual binding sites while DHPMs were created from hybrid-molecules occupying both binding sites. A huge library of 15, 63,764 publically available molecules were screened by CPMs and DHPMs. The screened hits obtained from both types of models were compared based on their Hashed binary molecular fingerprints and 4 point pharmacophore fingerprints using Tanimoto, Cosine, Dice and Tversky similarity matrices. Molecules screened by DHPM exhibited significant structural diversity, better binding strength and drug like properties as compared to the compounds screened by CPMs and the reported anti-mycobacterial molecules indicating the efficiency of DHPM to explore new chemical space for anti-TB drug discovery. The idea of DHPM can be applied for a wide range of mycobacterial or other pathogen targets to venture into unexplored chemical space.

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 severe limitations due to compliance and leads to emergence of drug resistant *Mycobacterium tuberculosis* (*Mtb*) strains [5]. Antimicrobial resistance (AMR) is one of the most serious global public health threats as, it compromises the successful prevention and treatment of deadly infectious diseases like tuberculosis [6]. 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*. In the last few decades, structure-based inhibitor design has become very popular in preclinical drug

development with the rapid advancements in the experimental techniques like X-ray crystallography or nuclear magnetic resonance, *in silico* tools and techniques as well as the computational power. More and more efficient algorithms are being implemented now a days for *ab initio* protein modeling, homology modeling, protein folding dynamics, molecular docking, pharmacophore modeling, virtual screening, quantitative structure activity relationship (QSAR) etc [7] [8] [9] [10] [11] [12] [13]. Several new strategies are also designed for repurposing existing drugs or find new hits [14]. The increasing number of experimental structures of drug targets deposited the protein data bank (PDB) and atomistic details of the binding modes of co-crystalized compounds/drugs/ natural ligands (substrates/cofactors) with these targets effectively aid to the structure-based lead generation and optimization. However, choosing the right molecular interaction features in the active sites of a drug target protein is the most important step of structure-based drug design [7, 10]. Most often, the interactions made by the natural substrates of the target protein/enzymes as reported in their static experimental structures are considered for designing new inhibitors. However, for target inhibition by competitive binding via specific interactions in the binding pocket, the static structure does not provide the correct picture of all potential and stable molecular interactions in the binding site. Also, this approach offer limited specificity and chemical diversity when cofactor/substrates are involved as these are common to both the host and the pathogen. Hence, utilization of additional information on adjacent binding pockets and their ligand binding potential along with the dynamics of the binding pockets might prove beneficial to design/screen novel chemotypes that can form specific stable interactions with multiple binding sites of the target offering better specificity and scope for introducing chemical diversity.

In this study we have considered the DapB enzyme of the Diaminopimelate (DAP) biosynthetic pathway from H37Rv strain of *Mtb* (*Mtb*-DapB), which is one of the validated drug targets [15] [16] as inhibition of this enzyme blocks the production of meso-diaminopimelate thus leading to inhibition of *de novo* lysine biosynthesis and peptidoglycan assembly (Fig. 1a), both of which are crucial for the survival of the pathogen [17]. Several groups made efforts to identify inhibitors of enzyme *Mtb*-DapB by exploring the potential of product analogues [18] 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 Ki values ranging from 7 to 48 μM [19]. 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 [20], [7], [8]. In this study, hybrid pharmacophore models representing multiple ligand binding regions of *Mtb*-DapB were designed based on the stable molecular interactions identified from the multiple MD trajectories of the target bound to the co-crystal ligand molecules as well as hybrids of

two such ligands bound to different binding sites (Fig. 1b). We comparatively investigated the efficacies of the dynamics-based hybrid pharmacophore models (DHPM) based on multiple binding sites and the conventional pharmacophore models obtained from a single binding site, to obtain novel, target specific and diverse chemotypes with higher binding affinities as compared to the existing ones.

Results And Discussion

Overall Structure of Mtb-DapB model systems

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. This protein basically binds two ligands, one is a cofactor either NADH or NADPH and a cyclic substrate. The ADP part of NADH/NADPH 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 unstable substrate dihydrodipicolinate which is located near the nicotinamide ring of NADH/NADPH placed in the N-terminal side of C1. Nine experimental structures of *Mtb*-DapB have been reported in PDB so far. Most of these crystal structures and literature suggest both NADH and NADPH can bind in the cofactor binding site of the n-terminal region, for the enzyme to function. Mostly a substrate mimetic competitive inhibitor 2,6-PDC (PDC) is bound at C1. In this study, we considered three crystal structures of Mtb-DapB holo form viz, 1P9L, 1C3V and 1YL7 as they are high resolution structures, the binding regions being occupied by NADH/NADPH and PDC so that all the ligand interaction features present in the binding sites can be thoroughly understood. Six different model systems were created from these PDB structures as narrated in the methodology section in order to represent various states of the protein in the cell, bound to one or more cofactors in presence and absence of the substrate molecule. With respect to the several stages in the reaction cycle, the geometric dynamic properties are quite likely to be different when the protein binds to different ligands and hence one needs to consider all these conformational states during ligand design. The hybrid molecules in **(D-NDPHyb)** and **(D-NAIHyb)** were designed in order to compare the stabilities of the interactions made individually by the ligands when they are free and when they are linked together with a linker. The stable interactions formed by the hybrid molecule can be mimicked to design inhibitors, which can compete for the binding sites of both the ligands of *Mtb*-DapB.

Comparative analysis of the structural dynamics of Mtb-DapB model systems

Since all our model systems are based on more or less similar initial structures, we wanted to study how binding of different ligands in different combinations affects the overall dynamics of Mtb-DapB and what conformation/s of which complex would be best to be used for drug design. Hence, various structural

properties were analysed from the MD trajectories of the six model systems using the 'Simulation event analysis' and 'Simulation interaction diagram' tools and we observed significant conformational variances in the binding sites of Mtb-DapB when bound to different ligands. Figure 4 shows the root mean squared deviations (RMSD) of the protein, root mean squared fluctuations (RMSF) of the Mtb-DapB residues RMSD, radius of gyration and solvent accessible surface areas (SASA) of the ligands in the six model systems.

All the trajectories were well equilibrated by the end of 200 ns as observed from the RMSD graph. Though the initial structures of all the model systems are very similar, the model systems with only NADPH/NADH as in **D-NDP** and **D-NAI** show high RMSD from the respective initial structures as compared to the ones that additionally bind PDC in C1. The model systems where both the ligands were linked together to form hybrid molecules **(D-NDPHyb and D-NAIHyb)** showed lesser RMSD from their respective initial structures as compared to the ones where the ligands are separate **(D-NDP-P and D-NAI-P)** suggesting relatively stable complexes with the hybrid molecules. Also the model systems binding NADH showed higher deviations from the initial structure except for **D-NAIHyb.** A quick look into the RMSD profiles of the ligand/s (NADPH/NADH and PDC considered together in case of **D-NDP-P and D-NAI-P**) with respect to themselves and with respect to the protein show very high fluctuations model systems, being comparatively lower in **D-NAI-P** as compared to **D-NAI.** These graphs suggest that NADH undergoes a drastic conformational change in **D-NAI-P** after 65 ns and then remains stable till 200 ns, while it almost detaches from the cavity after 130 ns in **D-NAI**. This observation is strengthened from radius of gyration and solvent accessible surface areas of the ligands throughout the simulations showing high fluctuations of NADH in **D-NAI**. As revealed from the C terminal L8 residues show higher (3.5 to 7.5 Å) fluctuations in all systems, which is however not a part of the binding pockets. In system **D-NAI**, the hinge region residue K136 shows high fluctuation.

Comparative analysis of stability of protein-ligand interactions in Mtb-**DapB model systems**

The simulation interaction diagram from Desmond gives the summary of all stable interactions between the active site residues and the ligands that last for more than 30% of the simulation. Figure 5 shows the fractions of different types of interactions made by the ligands with Mtb-DapB in the model systems and the stable interactions that lasted more than 30% of the simulation time for one system as an example, while **Fig. S1 and S2** show the same for all the six systems.

The phosphate groups of the ADP part of NADPH in **D-NDP-P** make stable H-bonds with positively charged residues K9 and R14. The phosphate groups also make stable water bridges with F52 and T53 with occupancies of 80% and 60% respectively. The –NH2 and -C = O groups of nicotinamide part make the most stable H-bonds with G75 (98%) and F105 (93%) as donor and acceptor respectively. The –OH group of the sugar moiety attached to the nicotinamide ring makes stable H-bond with T77. Intramolecular H-bonds were also formed between the sugar –OH and the adjacent phosphate groups. In absence of PDC as in **D-NDP**, the Adenine base of NADPH makes H-bonds with H54 (58%) and N61 (36%). The –OH and the phosphate groups attached to the sugar moiety adjacent to the adenine base

make water bridged interaction and H-bond with D33 (37%) and K9 (44%) respectively. The other phosphate group makes H-bond with F52 (42%). The sugar adjacent to the nicotinamide ring makes stable H-bonds with T33 (79%) and a water bridge interaction with T77 (64%). The H-bond interactions made by the –NH2 and -C = O groups of nicotinamide groups are same as those in **D-NDP-P**. The nicotinamide ring showed π-stacking with F17 as the latter is accessible due to absence of PDC. It was observed that many of the interactions formed by NADPH in D-NDP have occupancies < 50% whereas the interactions formed in D-NDP-P showed better occupancy suggesting relatively stable binding of NADPH to Mtb-DapB in presence of PDC. NADH shows relatively unstable binding with Mtb-DapB as compared to NADPH both in presence and absence of PDC. The adenine part of NADH in D-NAI-P makes H-bond interactions with S211 (40%) and a cation-π interaction with R214. The phosphate groups make interactions with positively charged R214 and K11. Nicotinamide –NH2 and –C = O groups still make H-bond and water bridge with G75 and F105 respectively. However, except the interactions of R214, none of the other interactions showed more than 50% occupancy suggesting a weaker binding. In absence of PDC in D-NAI, NADH showed no stable interaction with Mtb-DapB as it showed tendency to move out of the binding pocket. PDC showed many stable interactions with the C1 residues in presence of both NADPH and NADH. In D-NDP-P, PDC makes H-bond/electrostatic interactions with the hinge region residue K136 (two, 98% and 93%), R214 (95%), T143 (two, 94% and 93%) G142 (45%), T77 (40%), H133 (54%), water bridges with P103 and D138 and a π-stacking with H132 (56%). Similarly, in D-NAI-P, PDC makes H-bond/electrostatic interactions with K136 (two, 88% and 83%), H133 (69%), water bridges with D138, H132 and N104 and makes a π-stacking with H132 (40%). The number and stability of interactions made by PDC was observed to be higher in presence of NADPH than NADH. From these observations of interactions made by NADPH/NADH and PDC, we also get a hint that the strength of cofactor binding is higher in presence of the substrate. Another important observation is the interactions made by the hybrid molecules, phosphate groups of the hybrid molecule formed by linking NADPH and PDC (NDPHyb) made stable H-bond/electrostatic interactions with K509 (45%), R214 (two, 65% and 53%), the nicotinamide part made π-stacking with F52, and H-Bonds with F105 (45%) and A102 (60%). The PDC part of NDPHyb retained its stable interactions in the D-NDPHyb model system with K136 (99%), H133 (48%), T143 (two, 86% each), S141 (41%), G142 (75%) and water bridges with D138 (93%). It was interesting to observe that, when NADH was linked with PDC to form a hybrid molecule (NAIHyb), it formed more number of stable interactions with Mtb-DapB owing to its structural stability, which was not formed in **D-NAI-P** or **D-NAI**.

The sugar moiety attached to the adenine part of NAIHyb makes H-bond interactions with F52 (49%) and water bridge with A34 (42%). The phosphate groups make H-bond/electrostatic interactions with K11 (42%) and R214 (two, 88% and 87%) while the sugar attached to the nicotinamide ring makes H-bond with T77 (70%). The n-NH_2 and –OH groups of nicotinamide ring H-bond with T76 (61%), A102 (55%) and F105 (56%) respectively. The PDC part of NAIHyb retains its strong H-bond/electrostatic interactions with K136 (two, 99% each), T143 (two, 98%, 97%), G142 (87%), H133 (48%) and water bridges with D138 (97%), P103 (71%) and G78 (58%). Thus, the hybrid molecules form stable protein-ligand complexes by forming new additional interactions along with the *Mtb*-DapB binding site, especially in the C1 region.

From the above analyses it is clear that in all model systems, the interactions formed by the C1 residues were relatively more stable than the ones formed by other parts of the binding site. This observation was considered while generating hybrid pharmacophore models, which is discussed in the next section.

Dynamics Based Conventional And Hybrid 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. Each of the six MD trajectories was clustered using hierarchical clustering based on mutual RMSD to obtain representative conformations which have RMSD of least 3 Å with respect to each other. The **D-NDP-P, D-NDP, D-NDPHyb, D-NAI-P and D-NAIHyb** model systems gave rise to 3, 7, 3, 3 and 2 clusters respectively. DapB-NAI system was not considered for pharmacophore modelling as none of the interactions made by NADH in this model system showed > 40% occupancy. The cluster representatives were further verified if they are showing all the stable interactions as identified from the MD simulations. We observed that 2, 2, 3, 1 and 2 cluster representatives from the MD trajectories of **D-NDP-P, D-NDP, D-NDPHyb, D-NAI-P and D-NAIHyb** respectively were able to show all stable interactions of and then they were used to generate the N- and H-type pharmacophore models. Apart from the 4 representatives from **D-NDPHyb and D-NAIHyb**, the conventional models generated from the 2 representatives of D-NDP-P were combined to generate 2 H-type models. Thus, a total of 6 H-type and 6 N-type models were obtained as described in the methodology section.

Figure 6 shows one representative from each of the types of pharmacophore models and Tables S1, S2 and Fig.s S3 and S4 give the details of all the models belonging to the three categories. All the models consisted of 5 to 8 features. The conventional N-type models basically represented the most stable and energetically favoured interactions of NADPH/NADH and PDC with Mtb-DapB. The H-type models were designed to represent the stable interactions made by both PDC and the cofactor, excluding the ADP region. The reasons for excluding the ADP region are as follow, 1) ADP is a common metabolite in both human and Mtb, so mimicking its interaction features might cause specificity issue 2) combination of strong interaction features from both the cofactor and substrate might fetch chemically diverse molecules with good binding affinity and 3) interactions of the cofactors, substrate and PDC were relatively stable with the C1 residues (ADP region of the cofactors/hybrid molecules do not bind near C1). H-type pharmacophore models were obtained by two different ways. First, from the interactions of the Hybrid constructs (excluding the ADP part) as in **D-NDPHyb** and **D-NAIHyb** and second way was by merging the conventional models from NADPH/NADH and PDC of **D-NDP-P** followed by eliminating the features obtained from the ADP part of NADPH/NADH (Fig. 1). The ligand dataset was then screened by these N- and H-type models and molecules that matched at least 5 features were retained as hits. The unique consolidated list of compound hits screened by the N- and H-type models were named as N-set

and H-set respectively. N-Set contained 2884 molecules while the H-set contained 3707 molecules. **Fig. S4 and S5** show the pharmacophore models mapped to the respective top hits with best fitness scores.

Comparative Analysis Of N-set And H-set Compounds

As the aim of this study is to explore the hybrid pharmacophore model as an effective tool to screen new chemotypes with better affinities with DapB as compared to the conventional models, we compared different aspects of N-set and H-set compounds, such as their structural similarities, their binding energies with *Mtb*-DapB and drug like and ADMET properties.

Comparison of structural features

The structures of the H-set compounds were compared with those of the N-set compounds using different measures. We were also curious to see if these H-type models are able to screen compounds that represent new chemotypes as compared to the existing anti-TB molecules. So, we compared the H-set molecules with the anti-*Mtb* M-set molecules. The H-set compounds were chosen as the query library and compared against N- and M-sets as reference libraries. For each compound in the query library, the nearest neighbour in the reference libraries were obtained using fingerprint similarities based on different matrices. Four different types of binary hashed fingerprints such as linear, radial, MOLPRINT2D and atom triplets were calculated for the three libraries. The nearest neighbour similarities were obtained based on three different similarity matrices, viz., Tanimoto, Cosine, and Dice similarity matrices and plotted as histograms. **Fig. S6 and S7** show the similarity score distributions between H-set vs N-set and H-set vs M-set compounds respectively. Multiple fingerprints and similarity measures were used in order to ensure consistent results. Histograms in Fig. 7a show the distribution of the maximum similarity scores of the H-set compounds with respect to N- and M-set compounds based on the linear and radial fingerprints and Tanimoto and Cosine matrices. These graphs clearly indicate that more than 70% of the H-set compounds have scores below 0.6 with respect to the N-set compounds indicating that, the hybrid pharmacophore models screen structurally different molecules as compared to the conventional models. The striking dissimilarities (more than 98% of compounds with score < 0.5) between the H-set and M-set compounds as depicted in Fig. 7a indicated that the hybrid models screened novel chemical entities that are structurally different from the existing *Mtb* active molecules.

Comparison of Mtb-DapB binding affinities

In order to quantify the binding affinities of the H-set and N-set compounds with Mtb-DapB, all the 3707 H-set and 2884 N-set compounds were docked with each of the 10 representative structures (described in the ‘Dynamics based conventional and hybrid pharmacophore models’ section) used to generate the pharmacophore models. The best scoring pose for each compound among the 10 docking calculations was retained. This was followed by MMGBSA binding free energy calculations taking all these protein ligand complexes. Figure 7b shows the distributions of the XP docking scores and the

MMGBSA ΔG_{bind} ligand efficiency values of the top 1000 H-set and the top 1000 N-set compounds. The graphs show that, 72.3% of the top 1000 H-set compounds had docking scores below − 7, while only 34% of the N-set compounds have this range of XP-docking score. 20.1% of the top scoring 1000 H-set compounds had docking score <-8, while only 7.7% of the N-set compounds had this score. The MMGBSA ΔG_{bind} ligand efficiency values were obtained by normalizing the MMGBSA ΔG_{bind} by the number of heavy atoms present in the respective compounds. Figure 7b shows that 80.9% of the top 1000 H-set compounds showed ligand efficiency above 2.2, while only 40.4% of the top 1000 N-set compounds showed this range. These comparative graphs clearly indicate that the H-set compounds have better binding affinities with *Mtb*-DapB as compared to the N-set compounds. Examples of the best scoring H-set and N-set compounds have been shown in Fig. 7c.

Comparison of druglike properties

Various druglikeness parameters of the H-set and N-set compounds were calculated using QuickProp and the drug like properties of both sets were compared to each other and also with that of the existing anti-*Mtb* chemical space (M-set). Figure 8 shows the histograms of various druglikeness parameters of the H-, N-set and M-set compounds. The solubility, bioavailability and the druglikeness scores were found to follow strikingly different trends in case of N and H set compounds. **Fig. S8** shows distribution of different druglike properties of the H-set, N-set and M-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 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. 8) 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 as compared to the conventional models.

Conclusions

The present study reports a robust computational approach, wherein, six different model systems of *Mtb*-DapB binding different combinations of ligands were modelled and each of them were subjected to 200 ns molecular dynamics simulations. The structural and enthalpic stabilities of these models systems were monitored throughout the simulations and it was revealed that the hybrid ligands designed by linking two native ligands and substrate of *Mtb*-DapB are able to make highly stable non-covalent interactions in the binding pockets. These stable interactions formed by the hybrid ligands with two

adjacent binding site regions of *Mtb*-DapB were utilized to generate hybrid dynamics based pharmacophore models. The abilities of these hybrid models to screen molecules with new chemotypes, better binding affinities, and drug-like properties were comparatively assessed with that of the conventional models generated from the native ligands. Cheminformatics based structure comparison, docking scores, binding energies and ADMET properties of the molecules screened by the hybrid pharmacophore models were found to be more druglike, thus establishing the hybrid models as efficient tools to venture into novel anti-TB chemical space.

Material And Methods

Model Systems

Six different model systems were designed from three available experimental structures of Mtb-DapB for the study. The X-ray crystal structure 1C3V [21] was used to design three model systems, i.e., first one **(D-NDP-P)** with *Mtb*-DapB bound to NADPH (natural cofactor of Mtb-DapB) and 2,6-PDC (a mimetic of the natural substrate of *Mtb*-DapB) as in the reported structure, second one **(D-NDP)** bound to only NADPH and the third one **(D-NDPHyb)** bound to a hybrid molecule created by linking NADPH and 2,6-PDC with a simple n-propyl linker. Similarly, The x-ray crystal structure 1P9L [21] was used to generated two model systems, one **(D-NAI-P)** with Mtb-DapB bound to NADH (another natural ligand of Mtb-DapB) and 2,6-PDC as in the reported structure and the other model **(D-NAIHyb)** with a hybrid molecule created by linking NADH and 2,6-PDC with the n-propyl linker. The third crystal structure 1YL7 [22] was used as in the reported structure bound to only NADH **(D-NAI)**. Figure 1b depicts the model systems designed for this study.

Molecular Dynamics (md) Simulations

The Protein Preparation Wizard (PPW) module [23] of Schrödinger software package, version 2019-2 was used to pre-process the macromolecular structure downloaded from PDB. Missing hydrogens were added, water molecules beyond 5 Å of the active site were removed and appropriate bond orders were assigned to the structure. The residues/side chains unresolved in some of the crystal structures were repaired with prime [24] module in the PPW. The protonation states of the polar residues were optimized with the protassign module of PPW, which uses PROPKA to predict pKa values (pH 7.0 ± 2.0) and side chain functional group orientations. The structures were then subjected to restrained minimization (cutoff RMSD 0.3 Å) with impref to avoid steric clashes. These structures were further used for molecular dynamics (MD) simulations. MD simulations were carried out on the six model systems using the Desmond MD simulation package (release 2017) [25]. The OPLS_2005 [26] force field was employed for the protein-ligand complexes. Using the system builder tool of DESMOND the six model systems of Mtb-DapB were solvated in a cubical water box (TIP3P water model) keeping 10 Å buffer space in x, y and z dimensions. Each system was neutralized by adding appropriate counter ions and an ionic concentration of 0.15 M was maintained by adding Na^+ and Cl^- ions. The systems were minimized with 10000 steepest

descent steps followed by gradual heating from 0 to 300 K, under NVT ensemble. The systems were thermally relaxed before the production run using Nose-Hoover Chain thermostat method for 1 picosecond and 1 picosecond of pressure relaxation with Martyna-Tobias-Klein barostat method. Finally 0.2 microsecond (200 ns) production run under NPT ensemble was carried out for each system using a cutoff distance of 12 Å for non-bonded interactions. Simulation quality analysis, simulation event analysis, simulation interaction diagrams were used for trajectory analysis.

Md Trajectory Clustering And Generation Of E-pharmacophore Models

Hierarchical clustering was performed on each trajectory to sample representative structures with minimum RMSD of all the heavy atoms of the binding sites (residues within 5 Å of the bound ligands) of two representatives as 3 Å. The protein ligand interactions were analysed for each cluster representative and the ones with the most stable interaction features as identified from the MD trajectories were considered further for pharmacophore modelling using the e-Pharmacophore option from the Phase [27] module of Schrodinger Suite. To generate energy-based e-Pharmacophore models [28] from the given protein ligand complexes, Phase first estimates the Glide extra precision (XP) [29] energy terms for hydrophobic enclosure, hydrophobically packed correlated hydrogen bonds, electrostatic rewards, π-π stacking, cation-π, and other interactions. 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. Minimum inter-feature distance was maintained to be 2 Å, while minimum inter-feature distances between the same types of features were assigned to be 4 Å and the donors were presented as vectors. Receptor based excluded volume shells of 5 Å thickness were created using the Van der Waals radii of the receptor atoms. The scaling factor was assigned as 0.50 and receptor atoms within 2 Å of the ligand were ignored while defining the excluded volume shells as per the default settings. Conventional pharmacophore models were generated individually from NADPH, NADH and PDC from model systems **D-NDP-P, D-NDP** and **D-NAI-P.** Maximum number of features were assigned as 7 for these conventional models. None of the cluster representatives from the trajectory of **D-NAI** was used for pharmacophore model generation as none of the interactions formed between NADH and *Mtb*-DapB in this system were stable (occupancy < 40%). First type of hybrid pharmacophore models were generated directly from the hybrid molecules in model systems **D-NDPHyb** and **D-NAIHyb**. The maximum number of features were initially assigned as 10 for these hybrid models, but best 6 to 7 features were retained from the desired regions **(highlighted in yellow in** Fig. 1b**)** of the hybrid molecules, after discarding the features associated with the adenosine and phosphate regions. The second type of hybrid pharmacophore models with 6–7 features were obtained by first merging the conventional models from NADPH/NADH and PDC and then discarding the features associated with ADP regions of NADPH/NADH. The pharmacophore models obtained individually from 2, 6-PDC were having only 3 to 4 features, which would not confer specificity to the models. So they were not used further for ligand screening. The conventional pharmacophore models

obtained from NADPH/NADH individually were named as N-type models while the two types of hybrid pharmacophore models were named as H-type pharmacophore models.

Pharmacophore Screening

The ZINC natural product subsets with 132883 molecules [30] and Asinex screening library [31] consisting of 530881 molecules was selected for our study which included the “BioDesign” library of 175851 pharmacologically relevant natural product like compounds, “lead-like” library with 91473 compounds those have been screened against a panel of early ADMET tests (including DMSO and water solubility, PAMPA, PGP and CYP inhibition) and “Gold and Platinum Collections” with 263557 molecules having a high degree of drug-likeness, in accordance with Lipinski's Rule of 5. These molecules were subjected to preparation in LigPrep [32], generating their ionization states at pH 7.0 (± 2.0) using Epik ionizer and five lowest energy conformers were retained for each compound. Then, these molecules were screened against the N-type and H-type pharmacophore models using the “Ligand and Database Screening” option of Phase module of Schrodinger Suite [27]. 20 conformers were generated per molecule and the outputs were minimized before alignment against the pharmacophore models. We chose to consider the negative and acceptor features equivalent, the minimum number of sites the molecule must match was assigned to be 5. Among many conformers of a ligand, the one with the best fitness score (S) evaluated by a specific fitness function [33] was retained for each compound. The consolidated lists of all the compounds screened by the N- and H-type models were named as N-set and H-set respectively. Another dataset of 3412 non-redundant anti-*Mtb* compounds were prepared including the anti-*Mtb* compounds published by GSK [34], 347 *anti*-Mtb pathogen box compounds [35] and 1973 ChEMBL molecules [36] and 915 molecules form BioPhytMol [37] with < 2 μM activity on Mtb cells. This dataset was named as M-set. Various cheminformatics analyses were performed to compare the structural, physicochemical and ADMET properties of the N- and H-set set molecules with each other as well as with the M-set molecules. Mtb-DapB binding affinities of N- and H-set were estimated by molecular docking studies and compared.

Generation Of Molecular Fingerprints And Library Comparison

Different types of hashed binary molecular fingerprints such as liner, radial, MOLPRINT2D and 4 point 3D pharmacophoric fingerprints were generated for N-, H- and M-set molecules with the Schrodinger. Then, the H-set molecules were compared with the N-set and M-set compounds on the basis of the above molecular finger prints using Tanimoto, Cosine, Dice and Tversky similarity matrices. Max Similarity (MaxSim) scores were obtained for each molecule of the query set which represents the similarity score of a molecule with the nearest neighbour of the other set. Distribution of this MaxSim scores for the molecules of one set with respect to the other set gives an idea about the structural similarities between

the libraries. Various drug likeliness and ADMET properties were also calculated using QuikProp [35] for the three sets of molecules and were compared to each other.

Molecular Docking

The N- and H- set compounds were subjected to blind molecular docking with the cluster representatives used for pharmacophore model generation. As all the structures were prepared before MD simulations, these cluster representatives were directly used for grid generation. 'Receptor Grid Generation' module of Schrödinger was utilized to define interaction grids for molecular docking keeping the centroids of all residues within 5 Å of both the bound ligands as grid centres. The size of the interaction grid, which covered almost the whole protein including both the binding sites was fixed to 20 Å for inner box and 24 Å as outer box to facilitate a blind docking. Then, the N- and H-set compounds were subjected to docking calculations with the interaction grids of the selected cluster representatives using the Glide [29] module of Schrödinger software package first with standard precision (SP) followed by eXtra Precision (XP) modes. 5 best energy poses were generated for each compounds. OPLS_2005 force field was used for docking with all default parameters. The resultant complexes were further submitted for binding energy estimation, where Molecular Mechanics-Generalized Born Surface Area (MM/GBSA) based binding free energy (ΔG_{bind}) were computed for the complex using Prime module. The N- and H-type compounds were then compared based on their DapB binding affinities (Docking scores and ΔG_{bind}) as well as their potential to make interactions with key residues as identified from the MD trajectories. Figure 2 shows the overall workflow of the study.

List Of Abbreviations

Mycobacterium tuberculosis (*Mtb*)

Dynamics based hybrid pharmacophore models (DHPM)

Conventional pharmacophore models (CPM)

Quantitative structure activity relationship (QSAR)

Protein data bank (PDB)

Molecular dynamics (MD)

Root mean squared deviations (RMSD)

Root mean squared fluctuations (RMSF)

Solvent accessible surface areas (SASA)

Declarations

Availability of data and materials

Additional data is available through the following link:

Competing interests

The Authors declare none.

Funding

CC thanks Department of Science and Technology, New Delhi, India for financial support in the form of DST-INSPIRE Faculty award (IFA16-LSBM-170).

Authors' contributions

AB and CC conceived the project. CC performed calculations and analysed the results. AB and CC prepared the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

AB and CC thank Schrondinger. Inc. for providing short term licence for some modules.

References

1. Global Tuberculosis Report (2019) https://www.who.int/tb/global-report-2019. Accessed 4 Jun 2020
2. Balganesh TS, Alzari PM, Cole ST (2008) Rising standards for tuberculosis drug development. Trends Pharmacol Sci 29:576–581. https://doi.org/10.1016/j.tips.2008.08.001
3. Dutt AK, Stead W (1997) The treatment of tuberculosis. Disease-a-Month 43:247–274. https://doi.org/10.1016/S0011-5029(97)90023-7
4. Zumla A, Chakaya J, Centis R et al (2015) Tuberculosis treatment and management—an update on treatment regimens, trials, new drugs, and adjunct therapies. The Lancet Respiratory Medicine 3:220–234. https://doi.org/10.1016/S2213-2600(15)00063-6
5. Prasad R, Srivastava DK (2013) Multi drug and extensively drug-resistant TB (M/XDR-TB) management: Current issues. Clinical Epidemiology Global Health 1:124–128. https://doi.org/10.1016/j.cegh.2013.02.003

6. Passi A, Jolly B, Sharma T et al (2019) Chap. 8 - Data-Driven Systems Level Approaches for Drug Repurposing: Combating Drug Resistance in Priority Pathogens. In: Roy K (ed) In Silico Drug Design. Academic Press, pp 229–253
7. Choudhury C, Deva Priyakumar U, Sastry GN (2014) 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. https://doi.org/10.1016/j.jsb.2014.04.007
8. Choudhury C, Priyakumar UD, Sastry GN (2015) Dynamics Based Pharmacophore Models for Screening Potential Inhibitors of Mycobacterial Cyclopropane Synthase. J Chem Inf Model 55:848–860. https://doi.org/10.1021/ci500737b
9. CHOUDHURY C, PRIYAKUMAR UD, SASTRY GN (2016) Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors. J Chem Sci 128:719–732. https://doi.org/10.1007/s12039-016-1069-1
10. Choudhury C, Deva†Priyakumar U, Naraharı†Sastry G (2016) Structural and Functional Diversities of the Hexadecahydro-1H-cyclopenta[a]phenanthrene Framework, a Ubiquitous Scaffold in Steroidal Hormones. Mol Inf 35:145–157. https://doi.org/10.1002/minf.201600005
11. Gaur AS, Bhardwaj A, Sharma A et al (2017) Assessing therapeutic potential of molecules: molecular property diagnostic suite for tuberculosis $$(\mathbf{MPDS}^{\mathbf{TB}})$$(MPDSTB). J Chem Sci 129:515–531. https://doi.org/10.1007/s12039-017-1268-4
12. Kumar Srivastava H, Choudhury C, Narahari Sastry G (2012) The Efficacy of Conceptual DFT Descriptors and Docking Scores on the QSAR Models of HIV Protease Inhibitors. Med Chem 8:811–825
13. Kurumurthy C, Sambasiva Rao P, Veeraswamy B et al (2012) A facile and single pot strategy for the synthesis of novel naphthyridine derivatives under microwave irradiation conditions using ZnCl2 as catalyst, evaluation of AChE inhibitory activity, and molecular modeling studies. Med Chem Res 21:1785–1795. https://doi.org/10.1007/s00044-011-9695-0
14. Passi A, Rajput NK, Wild DJ, Bhardwaj A (2018) RepTB: a gene ontology based drug repurposing approach for tuberculosis. J Cheminform 10:24. https://doi.org/10.1186/s13321-018-0276-9
15. Cox RJ (1996) The DAP pathway to lysine as a target for antimicrobial agents. Nat Prod Rep 13:29–43. https://doi.org/10.1039/NP9961300029
16. Cox RJ, Sutherland A, Vederas JC (2000) Bacterial diaminopimelate metabolism as a target for antibiotic design. Bioorg Med Chem 8:843–871. https://doi.org/10.1016/S0968-0896(00)00044-4
17. Usha V, Lloyd AJ, Lovering AL, Besra GS (2012) Structure and function of Mycobacterium tuberculosis meso-diaminopimelic acid (DAP) biosynthetic enzymes. FEMS Microbiol Lett 330:10–16. https://doi.org/10.1111/j.1574-6968.2012.02527.x
18. Singh SP, Konwar BK, Bezbaruah RL, Bora TC (2013) Molecular docking and in silico studies on analogues of 2-methylheptyl isonicotinate with DHDPS enzyme of Mycobacterium tuberculosis. Med Chem Res 22:4755–4765. https://doi.org/10.1007/s00044-013-0488-5

19. Paiva AM, Vanderwall DE, Blanchard JS et al (2001) Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1545:67–77. https://doi.org/10.1016/S0167-4838(00)00262-4
20. Guterres H, Im W (2020) Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. J Chem Inf Model 60:2189–2198. https://doi.org/10.1021/acs.jcim.0c00057
21. Cirilli M, Zheng R, Scapin G, Blanchard JS (2003) The Three-Dimensional Structures of the Mycobacterium tuberculosis Dihydrodipicolinate Reductase – NADH – 2,6-PDC and – NADPH – 2,6-PDC Complexes. Structural Mutagenic Analysis of Relaxed Nucleotide Specificity Biochemistry 42:10644–10650. https://doi.org/10.1021/bi030044v
22. Janowski R, Kefala G, Weiss MS (2010) The structure of dihydrodipicolinate reductase (DapB) from Mycobacterium tuberculosis in three crystal forms. Acta Crystallographica Section D 66:61–72. https://doi.org/10.1107/S0907444909043960
23. Shivakumar D, Williams J, Wu Y et al (2010) Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. J Chem Theory Comput 6:1509–1519. https://doi.org/10.1021/ct900587b
24. Jacobson MP, Friesner RA, Xiang Z, Honig B (2002) On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. J Mol Biol 320:597–608. https://doi.org/10.1016/S0022-2836(02)00470-9
25. Shaw DE, Grossman JP, Bank JA et al (2014) Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE Press, New Orleans, Louisana, pp 41–53
26. Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110:1657–1666. https://doi.org/10.1021/ja00214a001
27. Dixon SL, Smondyrev AM, Knoll EH et al (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 20:647–671. https://doi.org/10.1007/s10822-006-9087-6
28. Salam NK, Nuti R, Sherman W (2009) Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. J Chem Inf Model 49:2356–2368. https://doi.org/10.1021/ci900212v
29. Friesner RA, Murphy RB, Repasky MP et al (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein – Ligand Complexes. J Med Chem 49:6177–6196. https://doi.org/10.1021/jm051256o
30. ZINC. http://zinc.docking.org/catalogs/subsets/np-derived/. Accessed 4 Jun 2020
31. Asinex.com – Asinex Focused Libraries, Screening compounds, Pre-plated Sets. http://www.asinex.com/. Accessed 4 Jun 2020

32. Shelley JC, Cholleti A, Frye LL et al (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des 21:681–691. https://doi.org/10.1007/s10822-007-9133-z
33. Dixon SL, Smondyrev AM, Rao SN (2006) PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. Chem Biol Drug Design 67:370–372. https://doi.org/10.1111/j.1747-0285.2006.00384.x
34. Ballell L, Bates RH, Young RJ et al (2013) Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. ChemMedChem 8:313–321. https://doi.org/10.1002/cmdc.201200428
35. The Pathogen Box | Medicines for Malaria Venture. https://www.mmv.org/mmv-open/pathogen-box. Accessed 4 Jun 2020
36. ChEMBL Database. https://www.ebi.ac.uk/chembl/. Accessed 4 Jun 2020
37. Sharma A, Dutta P, Sharma M et al (2014) BioPhytMol: a drug discovery community resource on anti-mycobacterial phytomolecules and plant extracts. J Cheminform 6:46. https://doi.org/10.1186/s13321-014-0046-2

Figures

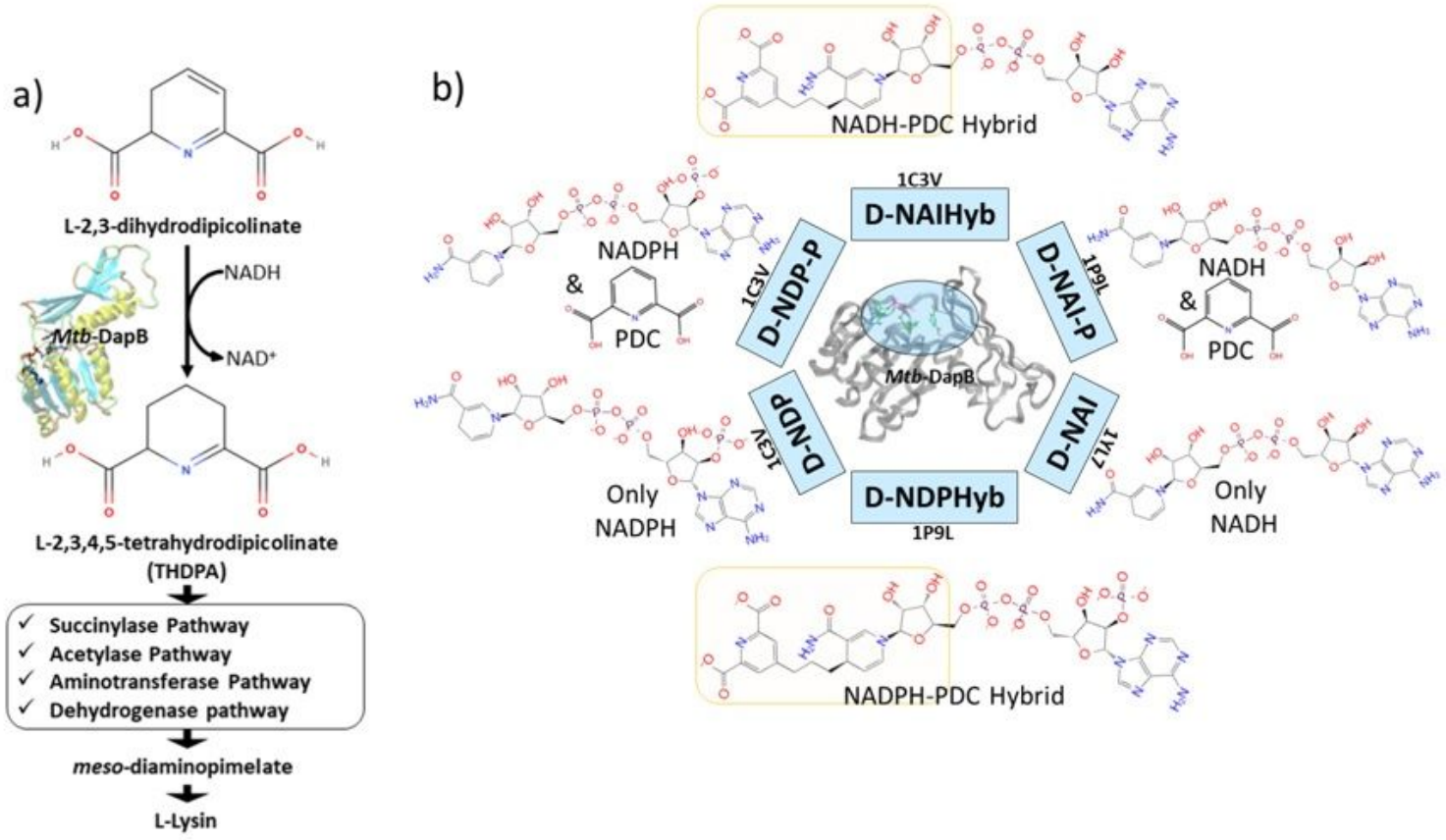


Figure 1

a) Schematic diagram showing the role of Mtb-DapB in de novo lysine biosynthesis and peptidoglycan assembly, b) Mtb-DapB model systems designed in the current study.

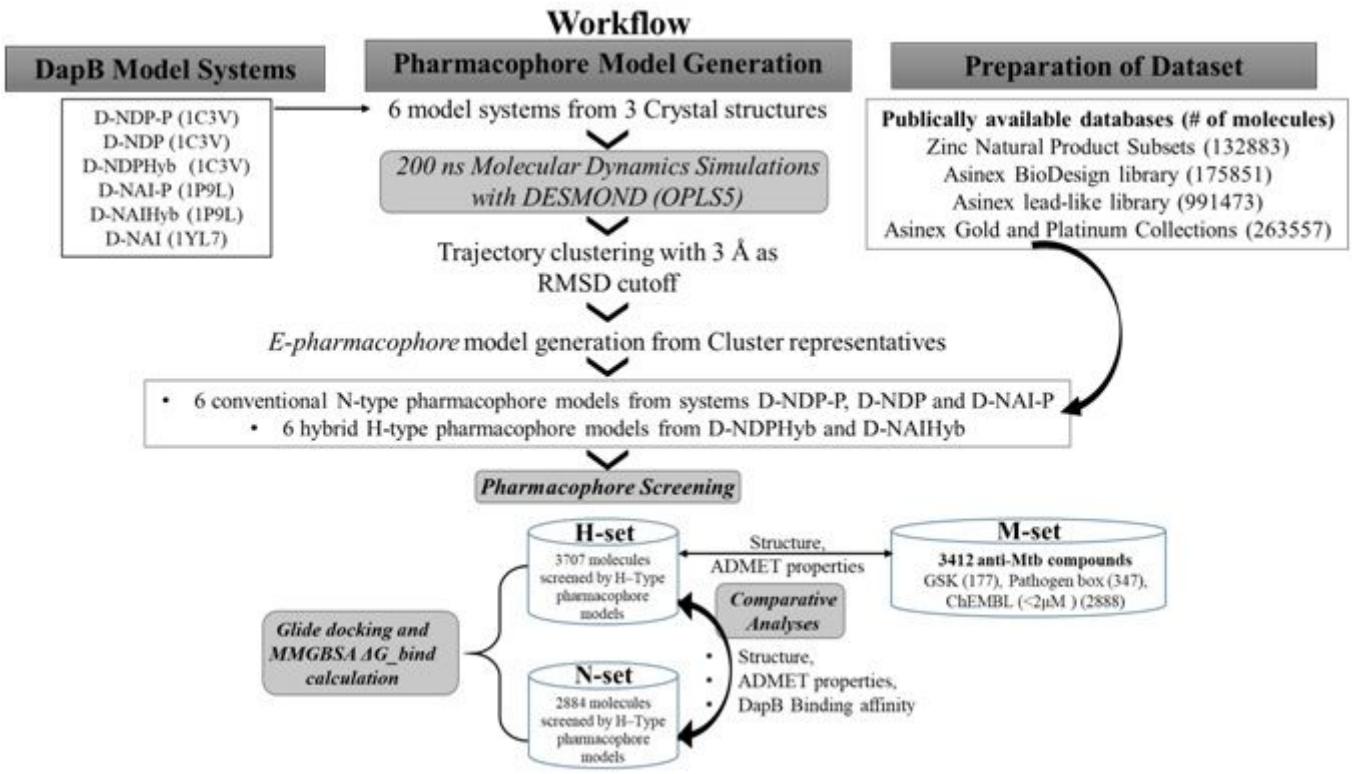


Figure 2

Overall workflow of the study.

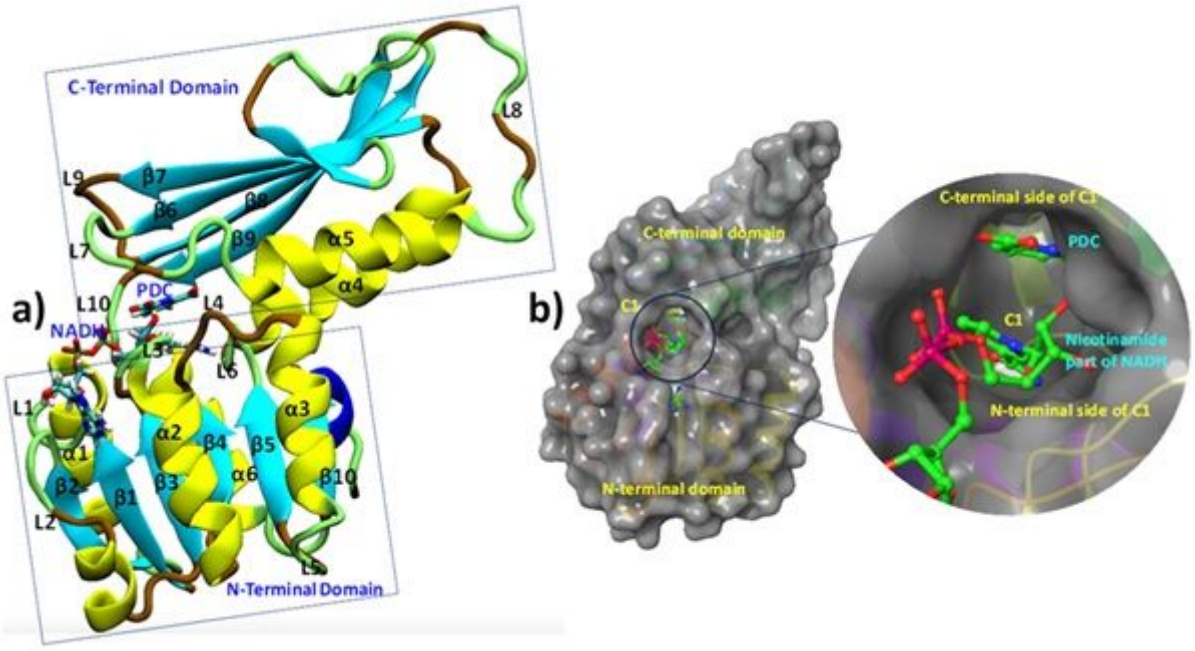


Figure 3

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

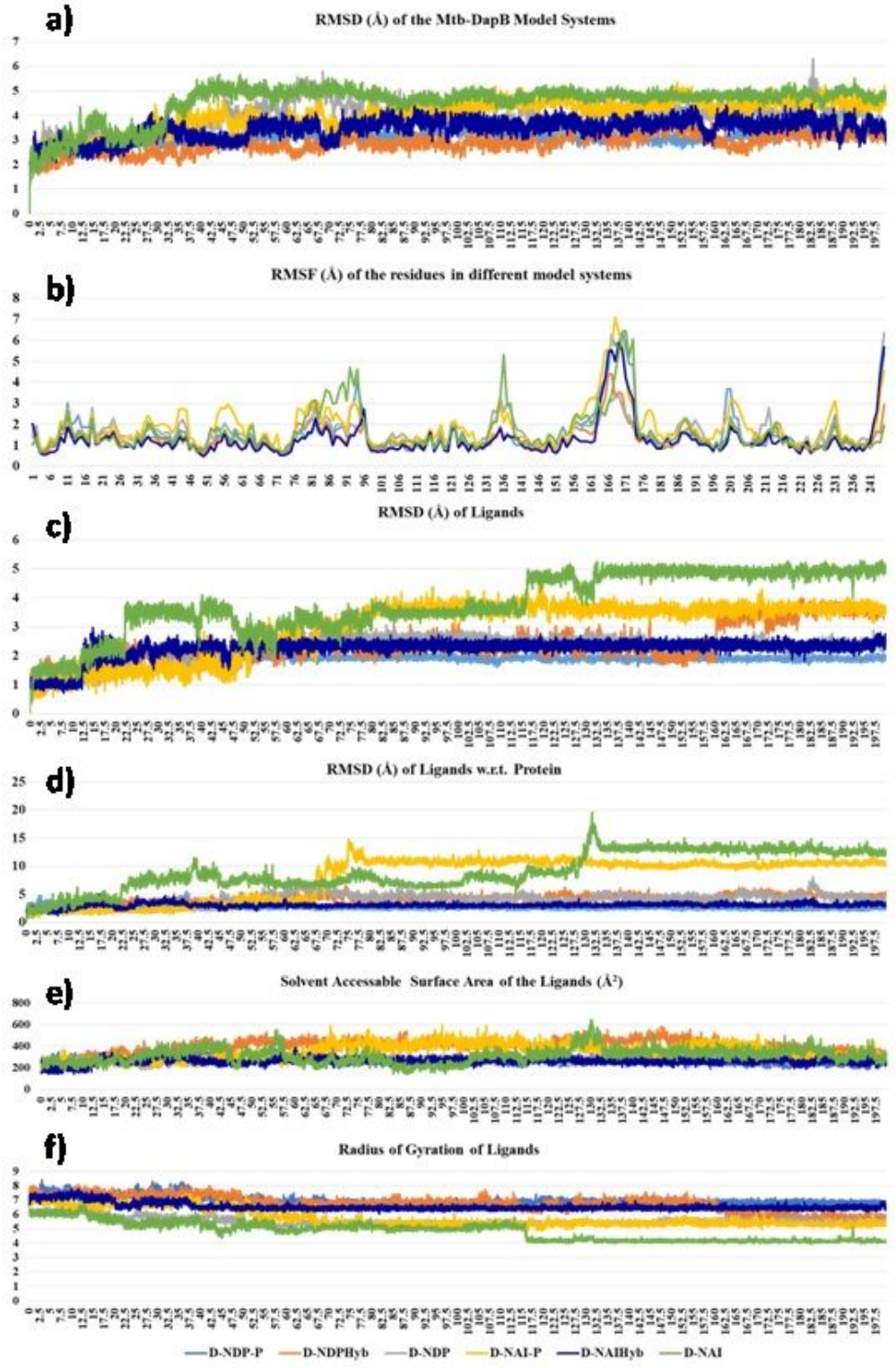


Figure 4

a) RMSD of the protein, b) RMSF of the Mtb-DapB residues, c) RMSD of the ligands, d) RMSD of the ligands wrt the protein, e) radius of gyration of the ligands f) SASA of the ligands in the six model systems.

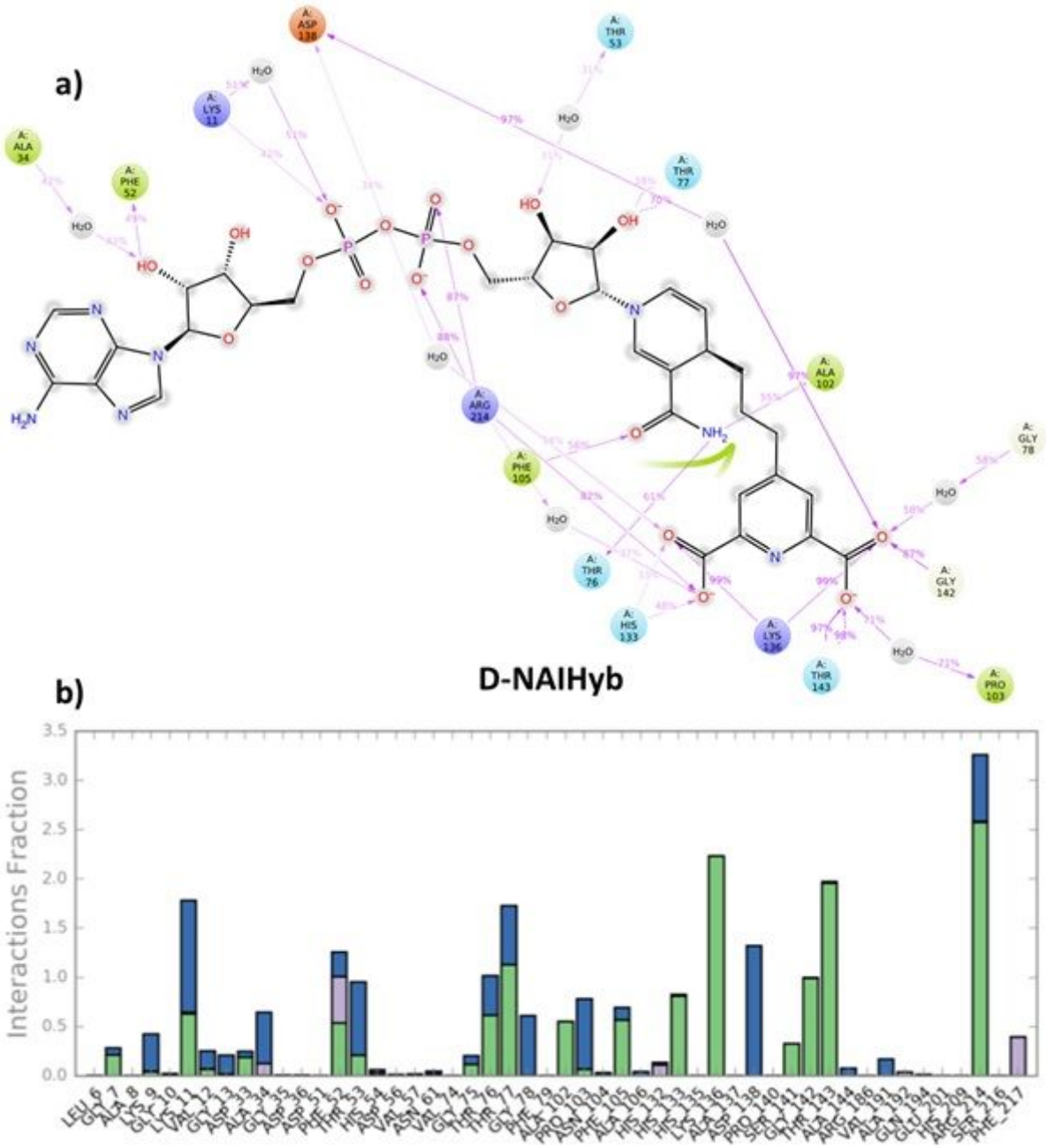
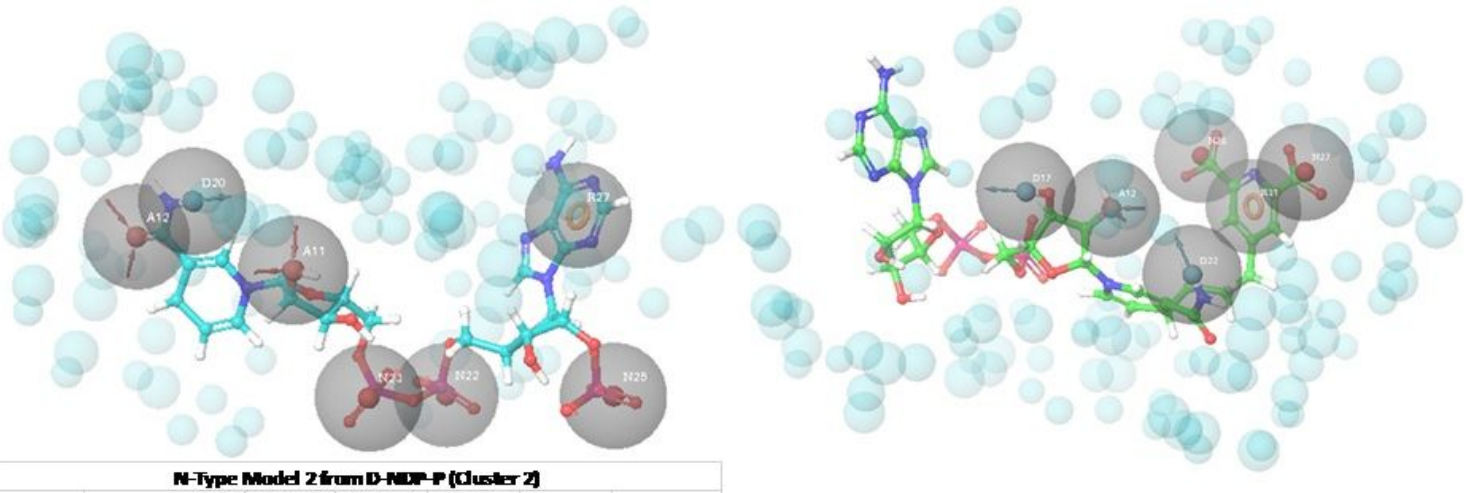


Figure 5

a) The stable interactions that lasted more than 30% of the simulation time, b) Fractions of different types of interactions made by the ligands with Mtb-DapB in the model system D-NAIHybas an example.



N-Type Model 2 from D-NDP-P (Cluster 2)

Rank	Feature_label	Score	X	Y	Z	source
1	A11	-3.20	30.46	9.00	-13.34	HBond
2	R27	-0.73	37.02	3.52	-5.87	Ring Chemscore
3	D20	-0.70	29.94	6.98	-17.89	HBond
4	N23	-0.69	27.73	7.65	-8.46	HBond
5	N22	-0.69	28.51	4.00	-6.07	HBond
6	N25	-0.62	32.57	5.64	-1.01	HBond
7	A12	-0.60	27.25	5.77	-18.99	HBond

H-Type Model 4 from D-NAIHyb (Cluster 2)

Rank	Feature_label	Score	X	Y	Z	source
1	A12	-1.60	2.38	-23.93	17.24	HBond
2	D17	-1.60	2.14	-26.73	14.84	HBond
3	N27	-1.25	1.90	-16.17	18.96	HBond
4	N28	-1.10	6.56	-19.80	19.04	HBond
5	D22	-0.70	-1.06	-21.93	19.50	HBond
6	R31	-0.59	3.72	-18.63	20.63	Ring Chemscore

Figure 6

Snapshots of one example from N-type and H-type pharmacophore models, with their respective feature tables.

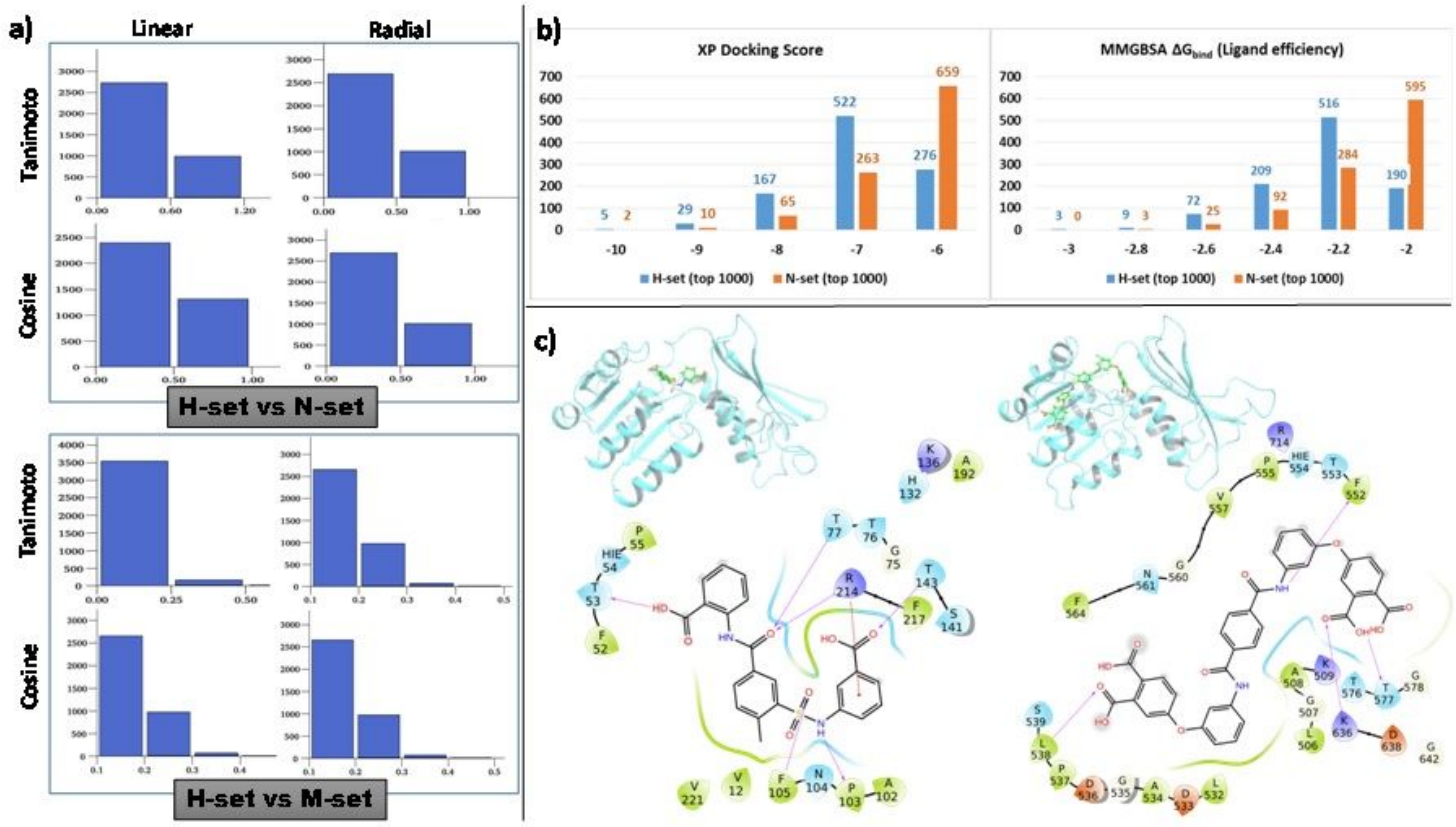


Figure 7

a) distribution of the maximum similarity scores of the H-set compounds with respect to N- and M-set compounds based on the linear and radial fingerprints and Tanimoto and Cosine matrices, b) Distributions of the XP docking scores and the MMGBSA ΔGbindligand efficiency values of the top 1000 H-set and the top 1000 N-set compounds, c) Examples of interactions of the best scoring H-set and N-set compounds bound to the respective binding sites.

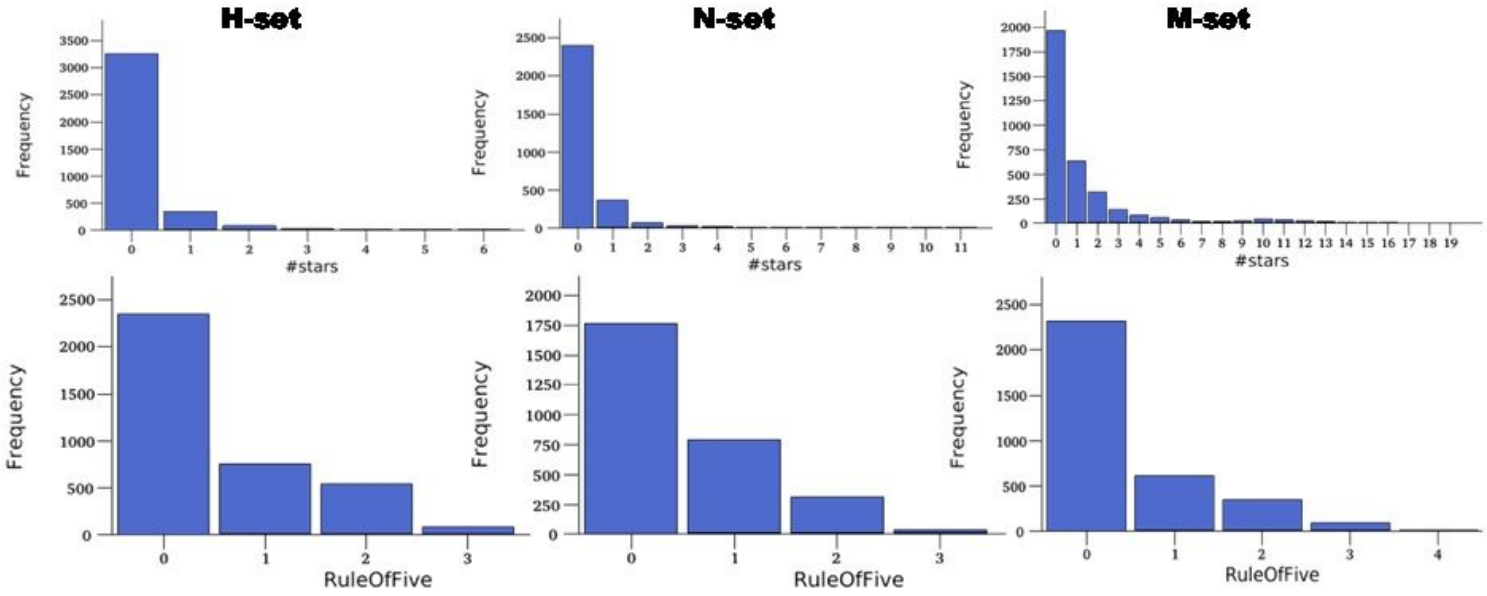


Figure 8

Histograms of important drug likeness parameters of the H-, N-set and M-set compounds.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarydata.docx