

Virtual Screening and Free Energy Estimation for Identifying Mycobacterium Tuberculosis Flavoenzyme DprE1 Inhibitors

Niranjan Kumar^{1†}, Rakesh Srivastava^{1†}, Amresh Prakash^{2*} and Andrew. M. Lynn^{1*}

¹School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi-110067, India

² Amity Institute of Integrative Sciences and Health, Amity University, Haryana, Gurgaon- 122413, India

†authors contributed equally

**Address for Correspondence*

Amresh Prakash, PhD
Assistant Professor
Amity Institute of Integrative Sciences and Health (AIISH),
Amity University Haryana, Gurgaon-122413, India
Email: amreshprakash@jnu.ac.in; aprakash@ggn.amity.edu

Prof. Andrew M. Lynn, Ph.D.
School of Computational & Integrative Sciences,
Jawaharlal Nehru University
New Delhi-110067
Email: andrew@jnu.ac.in

Abstract

In *Mycobacterium tuberculosis* (MTB), the cell wall synthesis flavoenzyme decaprenylphosphoryl- β -D-ribose 2'-epimerase (DprE1) plays a crucial role in host pathogenesis, virulence, lethality and survival under stress. The emergence of different variants of drug resistant MTB are one of the major threats worldwide which essentially requires more effective new drug molecules with no major side effects. Here, we used structure based virtual screening of bioactive molecules from ChEMBL database targeting DprE1, having bioactive 78,713 molecules known for anti-tuberculosis activity. An extensive molecular docking, binding affinity and pharmacokinetics profile filtering results in the selection four compounds, **C5** (ChEMBL2441313), **C6** (ChEMBL2338605), **C8** (ChEMBL441373) and **C10** (ChEMBL1607606) which may explore as potential drug candidates. The obtained results were validated with thirteen known DprE1 inhibitors. We further estimated the free-binding energy, solvation and entropy terms underlying the binding properties of DprE1-ligand interactions with the implication of MD simulation, MM-GBSA, MM-PBSA and MM-3D-RISM. Interestingly, we find that **C6** shows highest ΔG values (-41.28 ± 3.51 , -22.36 ± 3.17 , -10.33 ± 5.70 kcal mol⁻¹) in MM-GBSA, MM-PBSA and MM-3D-RISM assay, respectively. Whereas, the minimum ΔG scores (-35.31 ± 3.44 , -13.67 ± 2.65 , -3.40 ± 4.06 kcal mol⁻¹) observed for **CT319**, the inhibitor co-crystallized with DprE1. Collectively, the results demonstrated that hit-molecules **C5**, **C6**, **C8** and **C10** having better free binding energy and molecular affinity as compared to **CT319**. Thus, we proposed that selected compounds may be explored as lead molecules in MTB therapy.

Keywords: *Mycobacterium tuberculosis*; DprE1; Virtual screening; Bioavailability; MM-PBSA/GBSA; 3D-RISM.

1. Introduction

Mycobacterium tuberculosis (MTB) is a slow growing and widely spread pathogen, survive in both, intra-cellular and extracellular systems of patients, and infection may result in chronic and complex disease state. During the treatment, it can go to latency which revert to exponential growth on the immune defiance conditions of hosts [1, 2]. In recent years, WHO reports suggested that around 10.0 million (range, 9.0–11.1 million) individuals infected and 1.3 million (range, 1.2–1.4 million) people died from tuberculosis (TB) [1]. Moreover, the infection of MTB is one of the major causes of death worldwide, possessing the global health crisis, especially for the immunocompromised and HIV patients [3]. Although, the specific treatment may cure MTB, however, it requires multiple drug therapy for a longer period [1, 3]. Furthermore, the development of multi- and extensively-drug-resistant (MDR-TB and XDR-TB) MTB strains are the big challenges to control TB infections [4, 5]. In several conditions, it may turn into totally drug-resistant (TDR) tuberculosis which may worsen the condition of patients and therapy [2, 6]. Thus, the potential drug candidates, having minimal or no side effects are highly sought in MTB therapy [1, 2].

In recent years, several proteins involved in MTB survival and metabolism have been explored as potential drug targets and are progress in the drug development. During the evolution, mycobacteria have developed well-orchestrated and complex biosynthetic pathways to sustain a unique and thick cell wall which helps in maintaining the cellular integrity, survival under stress and dormancy, and eluding the host's immune systems. In MTB, the cell wall consists of the polymers of mycolyl-arabinogalactan-peptidoglycan, covalently connected with peptidoglycan and trehalose dimycolate that protects from stress, antibiotics and the hosts immune systems [7]. The flavoenzyme decaprenylphosphoryl- β -d-ribose 2'-epimerase (DprE1) involve in the biosynthesis of cell wall, plays critical role in formation of peptidoglycan-arbinogalactan-mycolic acid complex (PAM) and arabinogalactan and lipoarabinomannan (LAM) which are the essential building blocks and play crucial role in survival and

host pathogenesis, virulence, and lethality. DprE1 catalyses the first stage of epimerization reaction especially in the presence of FAD, it oxidizes C2' hydroxyl site of DPR to produce the keto intermediary decaprenyl-2'-keto-D-arabinose(DPX) and then DPA is formed by using decaprenyl-phosphoryl-D-2-keto-erythro-pentose reductase (DprE2) and reduced form of nicotinamide adenine dinucleotide (NADH) as a cofactor [8-10]. Thus, the catalytic activity of DprE1 is one of the potential drug targets in the development of tuberculosis therapy [2, 4, 7]. Recently, the benzothiazinones (BTZs) derivatives have shown higher potency for inhibition of DprE1, and efficacy against XDR and MDR mycobacterium clinical isolates.

To improve the pharmacological properties of the compounds, chemical scaffold piperazine was added to BTZ. Further, the lead optimization of PBTZ derivatives results in the discovery of more potent compounds which are currently in clinical trials [5, 8, 11]. In this view, several structurally distinct chemical scaffolds are in drug screening as DprE1 inhibitors. Broadly, these inhibitors can be categorized as covalent or noncovalent, distinctly involved in interaction at the catalytic domain of DprE1 [8, 11]. To elucidate the action and interaction of BTZs compounds, Batt et al., solved the X-ray crystal structure of DPrE1 in both, ligand free and bound form. He found that the structure of DprE1 consists of two functional domains, FAD binding domain and substrate binding domain. The co-factor was buried deeply in highly conserved FAD domain. The substrate binding extended for FAD, decorated largely with antiparallel β -strands (β_{10-16}) and included disordered loops at surface which govern the wide and open active site. The nitroaromatic inhibitors (e.g., BTZ, VI-9376, nitroimidazole 377790) possesses nitro moiety which involved in covalent interaction at C387, whereas, the noncovalent inhibitors (e.g., TCA1, 1,4-azaindoles, pyrazolopyridones, 4-aminoquinolone piperidine amides, Ty38c) potentially inhibit the enzymatic function of DprE1 showed that hydrophobic, electrostatic, and van der Waals interactions are critical for the spatial stability of inhibitors at the active site of DprE1 [5, 11]. Thus, the exploration of

crystal structure of DprE1 has been largely facilitated the drug discovery efforts to tend the molecules effective against MDR and XDR strains [2, 5, 8, 11].

Recent studies on the development of DprE1 inhibitors suggested a major contribution of molecular modelling, high throughput screening, docking, functional genomics and proteomics in paradigm of identifying novel chemical scaffolds as potential molecules for TB chemotherapy [5, 8, 11]. Although, molecular docking programs provide the description of protein-ligand interactions. However, a better understanding of protein-ligands interactions requires an accurate description of the spatial orientation of ligands at the active site of protein, conformational dynamics of protein and active sites residues, interaction energy and molecular stability [12-14]. In this view, MD simulation is an efficient and well-established computational method which mimics the flexible nature of bio-molecules, protein conformational changes, protein-ligand interactions, structural perturbation and provide more realistic picture with atomic details in reference to time [4, 15, 16]. Moreover, the free binding energy estimation, effect of solvation and thermodynamic integration is the central focus to understand the molecular interactions which can be well achieved by the implication of MM-GBSA, MM-PBSA and MM-3D-RISM using the trajectories obtained from MD simulation [12, 17-19].

In this context, we employed the structure based virtual screening for identification of promising chemical entities as DprE1 inhibitors from the ChEMBL database. We find that 78,713 small molecules at ChEMBL database suggested for the anti-mycobacterial activity. The three steps molecular docking and binding affinity estimation process lead to the selection of 10 hit-molecules. Similar procedures were applied on the selected 13 DprE1 inhibitors for the comparison of results with hit-molecules. Multiple MD simulations were performed on the DprE1 complex with hit-molecules and inhibitor (**CT319**) and the spatial stability of ligand molecules at active site of protein was estimated in terms of binding free energy using MM/PBSA/GBSA, and MM-3D-RISM [15, 17, 18]. The extensive evaluation of pharmacokinetic profile and drug-likeness properties analyses suggested that four chemical entities,

compounds **C5** (ChEMBL2441313), **C6** (ChEMBL2338605), **C8** (ChEMBL441373) and **C10** (ChEMBL1607606) may be explored as potential lead molecules for the development of promising DprE1 inhibitors in MTB therapy.

2. Materials and Methods

2.1. Protein preparation

The X-ray structure of DprE1 with inhibitor CT319 and cofactor FAD (PDB ID: 4FDO) was taken from the protein data bank (www.rcsb.org) [20]. The structure of DprE1 consist of two domains, the FAD binding domain comprised with α/β folds (residues 7–196, 413–461) and another domain, substrate binding includes extended conformation and antiparallel β -sheets (residues 197–412). In the crystal structure, the spatial orientation of FAD-binding domain and residues involved in interactions were highly conserved and critical, the cofactor is deeply buried, with the isoalloxazine at the interface to the substrate-binding domain [11]. And, the substrate-binding domain orientated towards the interface of flavin binding at centre. Thus, to prepare the protein files for molecular docking studies, the bound complex of protein with cofactor FAD was used. The other heteroatoms, co-crystallised inhibitor (**CT319**) and water molecules were removed. The structural regions lacking for low electron density were prepared, using Chimera tools [21]. Protein preparation wizard of Glide was used to assign hydrogen atoms and examine the structural correctness [22]. Finally, the optimized coordinates of DprE1 with FAD was used to carry out molecular docking and virtual screening against the selected antimycobacterial compounds from ChEMBL database.

2.2 Ligand preparation

Here, we used the anti-tuberculosis compounds taken from the publicly available chemical compounds database, ChEMBL. It consists of 2,101,843 compounds, out of these 78,713 compounds were bioactive molecules, having anti-tuberculosis activity downloaded from the ChEMBL database [23]. After sorting

of compounds based on the repetitive entries, 30,789 ChEMBL compounds were found unique which were used for the ligand preparation. The SMILES (simplified molecular-input line-entry system) strings formats of compounds were converted to 3D SDF format, missing hydrogen atoms were added, and the structures were optimized using CORINA v2.64 software package [24]. The module, Ligprep of Schrodinger suite 2017-3 used to generate compounds with low energy 3D structures [25]. The ionization/ tautomeric states of the selected compounds were taken care of by Epik parameters. The compounds chirality was taken from the original state. All the conformations were minimized and produced at a maximum of 32 conformations per ligand using the OPLS-2005 force field at a pH 7 ± 2 [26-28].

2.3 Virtual Screening and Scoring

The structure based virtual screening of compounds against DprE1 was performed using Glide, Schrodinger, LLC [29, 30]. The grid box define over the active site of DprE1, having outer box size X= 30, Y= 30, Z= 30 with grid center, X=40.1971, Y= 16.829, and Z= 9.172 [11]. The high throughput virtual screening involved three step filtering processes (i) selection of top ten percent ligand molecules using standard precision, (ii) then, docking of molecules by the extra precision mode of Glide (Glide XP) which allow the flexibility of ligands and (iii) the best-docked compounds were chosen using a Glide Emodel energy, Glide energy and Glide score function. The Glide Emodel includes a combination of Coulombic and van der Waals interaction energy, Glide score and strain energy of ligands which were used to selected lowest energy docked complexes on which the post-docking analyses were performed. Further analysis involved the re-scoring of selected docked complex using X-score v1.2.1 [31] and the protein-ligand molecular interactions were examined using Ligplot [32] and Discovery Studio Visualizer (Accelrys, San Diego, CA, USA).

2.4 ADME studies

The bioavailability of selected lead molecules, ADME (adsorption, distribution, metabolism and excretion) properties were calculated using module Qikprop v5.7 of Schrodinger 2018-3. ADME descriptors includes, central nervous system (CNS), molecular weight (MW), prediction of octanol/water partition coefficient (QPlogPo/w), aqueous solubility (QPlogS), IC50 value for blockage of HERG K⁺ channels (QPlogHERG), gut blood barrier (QPPCaco), brain/blood partition coefficient (QPlogBB), binding to human serum albumin (QPlogKhsa), Lipinski's rule of five (RO5) and percentage of human oral absorption (% of Human oral absorption).

2.5 In-silico drug-likeness and toxicity prediction

The molecular properties predictor tool, OSIRIS was used for the prediction of side effect risks of the hit-molecules, such as mutagenicity, tumorigenicity, irritant and reproductive effects. It also calculates the drug-relevant properties: cLogP, solubility (LogS), molecular weight (MW) and based on overall drug-score suggested the drug-likeness properties of molecules [33].

2.6 Molecular dynamics (MD) simulation

MD simulation was performed on the coordinates of DprE1 and DprE1-ligand complexes, using Amber16 biosimulation package. The force field ff14SB with TIP3P water model was used for the solvation of prepared systems. The charges, parameters and force field for cofactor (FAD) and ligands were defined by AM1-BCC charges and force field GAFF, using Antechamber tool. Here, six independent MD runs were carried out for the prepared systems, DprE1-FAD, DprE1-FAD-CT319, DprE1-FAD-ChEMBL1607606, DprE1-FAD-ChEMBL2338605, DprE1-FAD-ChEMBL2441313 and DprE1-FAD-ChEMBL441373 complexes. And, the systems were prepared using tleap tool of Amber16 with buffer distance (12 Å) in the octahedral simulation box. To neutralize the system 0.15 M counter ions (Na⁺ and Cl⁻) were added [34]. Bonds involving hydrogens were treated with SHAKE algorithm and the long-range electrostatic forces were handled using Particle mesh Ewald summation. During the

simulation, Berendsen's barostat and Langevin thermostat were used to maintain the Pressure and temperature, respectively. The energy minimization processes involved two phases. First phase included 3000 minimization steps, which involved 2500 steps of steepest gradient and remaining 500 steps of conjugate gradient algorithm. The solute atoms were restrained ($100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) and the only movements of water molecules and counter ions were allowed. The second phase of minimization included 5000 minimization steps (steepest gradient: 4500 steps and conjugate gradient: 500 used) without restraints on any atom. Minimization step followed by heating equilibration of system from 0 to 298 K with a time step of 1fs for 30 ps and consecutive equilibration run of 100 ps using time step of 2 fs with NPT ensemble. Using pmemd.cuda, the production run was performed on NPT ensemble for period of 100 ns and the time step was set to 2 fs. All files, trajectories, velocity and energy were saved at a gap of every 10 ps. The simulation trajectories were analysed using cpptraj tool available in Amber16.

2.7 Binding free energy calculation

Free energy change of a protein-ligand binding can be represented as follows:

$$\Delta G = \langle G_{RL} \rangle - \langle G_R \rangle - \langle G_L \rangle \quad \text{eq. (2)}$$

where, ΔG , G_{RL} , G_R and G_L represent the binding free energy of protein-ligand system, free energy of protein complexed with ligand, free energy of protein and free energy of ligand, respectively. The angular brackets represent ensemble average. Neglecting the entropy change of protein and ligand as a result of binding, equation (eq.) 2 can approximately be written as:

$$\Delta G = \Delta E + \Delta G_{SOL} \quad \text{eq. (3)}$$

where, ΔE is the interaction energy change (gas-phase) upon ligand binding. ΔG_{SOL} is the solvation free energy change on ligand binding. Here, ΔE can be computed using molecular-mechanics force field and the second term, ΔG_{SOL} can be estimated with the help of a proper solvation model.

Solvation models can be categorized in one of the two classes, implicit solvation model and explicit solvation model. Implicit solvation models consider the solvent molecules and dissolved salt ions as a mean field dielectric continuum. In contrary to implicit solvation models, explicit solvation models define solvent species at atomistic detail. Generalized Born Surface Area (GBSA), Poisson-Boltzmann Surface area (PBSA) are the two most common implicit solvation models used for solvation free energy calculation. Whereas, using the first principles 3D-RISM-KH (three-dimensional reference interaction site model with Kovalenko–Hirata) provides a 3D maps of solvation structure, thermodynamics and, more accurately predicting the parameters accounts for the ligands binding interactions and affinities.

2.7.1 Generalized Born Surface Area (GBSA)

In GBSA approach, the solvation free energy of a solute molecule, G_{SOL} is calculated in two parts: polar or electrostatics ($G_{\text{SOL-GB}}$) and nonpolar or non-electrostatics (G_{SURF}) ($G_{\text{SOL}} = G_{\text{SOL-GB}} + G_{\text{SURF}}$).

$G_{\text{SOL-GB}}$ is estimated using the following generalized Born expression:

$$G_{\text{SOL-GB}} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{\text{GB}}(r_{ij}, R_i, R_j)} \left(1 - \frac{\exp(-\kappa f_{\text{GB}})}{\epsilon} \right) \quad (\text{eq. 4})$$

where q_i , q_j represent charges on solute atoms i , j respectively and r_{ij} is the distance between them, R_i represents effective Born radii (estimated using van der Waals radius and burial of atom),

$$f_{\text{GB}} = \sqrt{r_{ij}^2 + R_i R_j \exp(-r_{ij}^2 / 4R_i R_j)}$$

and the summation runs over all pairs of atoms in the solute molecule. The screening effect produced by the monovalent salt ions is incorporated in eq. 4 through the Debye-Huckel screening parameter κ . The nonpolar or non-electrostatics contribution to solvation free energy (also known as cavitation term), G_{SURF} is determined on the basis of solvent accessible surface area (SASA) of molecules.

2.7.2 Poisson Boltzmann Surface Area (PBSA)

In PBSA model, the solute is represented in atomic detail with molecular mechanics force field and the solvent molecules along with dissolved electrolytes is represented as a dielectric continuum. This approach considers the solute molecule as a dielectric object whose shape is determined by the atomic coordinates and their radii. Electric charges present on atoms of solute molecule produce electric field and in response the solvent also produce a reaction field. The electrostatic potential $\phi(\underline{r})$ at a point satisfies the Poisson-Boltzmann (PB) equation and can be computed by solving it:

$$\nabla \cdot [\epsilon(\underline{r}) \nabla \phi(\underline{r})] = -4\pi\rho(\underline{r}) - 4\pi \sum_i z_i c_i \exp(-z_i \phi(\underline{r})/kT) \quad (\text{eq. 5})$$

Where, $\epsilon(\underline{r})$ is the dielectric constant, $\phi(\underline{r})$ is the electrostatic potential, $\rho(\underline{r})$ is the solute charge density at position \underline{r} , z_i is the charge on ion i , c_i is the bulk number density of ion i , k is the Boltzmann constant, and T is the absolute temperature; the summation is over all different ion types. when the solute does not carry a high charge, the second term in eq. 5 can be linearized and it results into linearized PB equation. In Amber PB equation is solved numerically for solutes of arbitrary shape which gives electrostatic potential $\phi(\underline{r})$ at each point of the system. Once we know the electrostatic potential $\phi(\underline{r})$ at each point, we can calculate the polar solvation free energy ($G_{\text{SOL-PB}}$) of the solute by multiplying each solute charge q_i by electrostatic potential $\phi(\underline{r}_i)$ at that point. Here, the nonpolar part of solvation free energy G_{NPOL} is also calculated through SASA of solute, as described in GBSA analyses.

2.7.3 Three-dimensional reference interaction site model with KH closure (3D-RISM-KH)

3DRISM is a semi analytical theory based on statistical mechanical Ornstein-Zernike (OZ) equation which is a contrary to MM/GBSA and MM/PBSA, considers molecular structure of solvent and salt ions. OZ equation splits the total correlation between two molecules into direct correlation between them and indirect correlation, which comes from other particles present in the system. In this theory, the molecular interactions are converted to sum of site-site interactions where atomic centres are taken as sites. The

three-dimensional solvent distribution functions are obtained from the solution of following 3DRISM integral equation:

$$h_{\gamma}(\mathbf{r}) = \sum_{\alpha} \int d\mathbf{r}' c_{\alpha}(\mathbf{r} - \mathbf{r}') \chi_{\alpha\gamma}(\mathbf{r}') \quad (\text{eq. 6})$$

Where, $h_{\gamma}(\mathbf{r}) = g_{\gamma}(\mathbf{r}) - 1$ and $c_{\gamma}(\mathbf{r})$ are the total and direct correlations of solvent in 3D and summation is taken over all interaction sites of all solvent species. The susceptibility function $\chi_{\alpha\gamma}(\mathbf{r}')$ for solvent was calculated using dielectrically consistent reference interaction site model (DRISM) theory and used as input to 3DRISM calculation. As eq. 6 involves two variables; hand c, therefore, we need another closure relation to solve it. Here, we have used Kovalenko-Hirata (KH) closure which is as following:

$$\begin{aligned} g(\mathbf{r}) &= \exp(-\beta u_{\gamma}(\mathbf{r}) + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r})) \text{ for } g_{\gamma}(\mathbf{r}) \leq 1 \\ \text{and } g(\mathbf{r}) &= 1 - \beta u_{\gamma}(\mathbf{r}) + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r}) \text{ for } g_{\gamma}(\mathbf{r}) > 1 \end{aligned} \quad (\text{eq. 7})$$

The one-dimensional site-site solvent susceptibility of solvent is defined in two parts as follows:

$$\chi_{\alpha\gamma}(\mathbf{r}) = \omega_{\alpha\gamma}(\mathbf{r}) + \rho_{\alpha} h_{\alpha\gamma}(\mathbf{r})$$

Where, the intramolecular correlation function $\omega_{\alpha\gamma}(\mathbf{r})$ incorporates the molecular geometry of solvent and $h_{\alpha\gamma}(\mathbf{r})$ is the total correlation between solvent sites α and γ .

In Amber, eqs. (6) and (7) are solved numerically to obtain three-dimensional solvent distribution functions around a fixed solute geometry and solvation free energy of the solute is calculated using the following equation which is an extension of Singer-Chandler formula:

$$G_{\text{SOL}} = \sum_i \frac{\rho_i}{\beta} \int_0^{\infty} d\mathbf{r} \left[-c_i(\mathbf{r}) - \frac{1}{2} h_i(\mathbf{r}) c_i(\mathbf{r}) + \frac{1}{2} \{h_i(\mathbf{r})^2 \Theta(-h_i(\mathbf{r}))\} \right] \quad \text{eq. (8)}$$

Where, Θ is the heaviside function and the summation runs over all the solvent sites i . The nonpolar part of $G_{\text{SOL-NPOL}}$ is calculated by assigning all solute charges zero. Further, the solvation free energy G_{SOL} can also be decomposed into energetic and entropic components ($\Delta G_{\text{SOL-E}}$ and $-\Delta G_{\text{SOL-TS}}$) using temperature derivatives.

3. Results and Discussion

The drug development process involves several expansive steps and complex strategies. Recent advancement in the computational modelling techniques, molecular docking, high-throughput virtual screening, pharmacokinetic profile (ADME), toxicity and bioavailability analyses of the molecules have been perceived as well-established techniques to accelerate the drug development processes [13, 31, 35-37]. Further, the integration of MD simulation and estimation of free-binding energy provide an accuracy on the spatial fitting, interaction stability and binding affinity of ligands at the active site of protein [15, 17-19, 38]. Herein, we systematically utilized the structure based virtual screening of compounds from ChEMBL database, having bioactive 78,713 molecules known for anti-tuberculosis activity (**Figure 1**). The initial sorting of molecules leads to selection of 30,789 molecules which were subjected for molecular docking against the protein DprE1, the oxidoreductase enzyme involved in cell wall synthesis of MTB. The extensive ADME, toxicity and pharmacokinetic profile analyses were performed on hit-molecules which results in the selection of four ChEMBL compounds as potential lead molecules for DprE1 inhibitors. To improvise the molecular docking results, multiple MD simulations carried out on DprE1, DprE1-CT319 and DprE1- -hit molecules (**C5, C6, C8, C10**) complexes. Typically, MD simulation deciphers the structural stability of protein-ligand interaction, conformational orientation, stability and molecular interactions of ligands at active site [11, 22, 39]. Moreover, the obtained MD trajectories utilized to calculate MM/PBSA, MM/GBSA and 3D-RISM-KH, which provide a robust estimation of free-binding energy, contacts and effect of solvent underlying the binding affinity of ligand molecules [12, 17, 18].

3.1 Virtual screening and docking analysis against DprE1

The structure based virtual screening was performed on 30,789 small molecules taken from ChEMBL database, having anti-tuberculosis biological activity. Glide based molecular docking involves various filtering steps for the high throughput virtual screening (HTVS) [29]. During the initial step, docking

leads to the selection of 3,078 compounds, top scored 10 % compounds. These top scored 10 % compounds are subjected for standard precision (SP) docking which results in the selection of 307 compounds. The extra precision (XP) filtering is applied on another top scored 10 % (307 molecules) hit-molecules. Finally, the best scored top 10 hit-molecules are selected for the comparative studies with known DprE1 inhibitors and drug molecules taken from the recent literatures [5, 8, 11]. The 2D interactions of top 10 hit-molecules and 13 inhibitors are shown in **Supplementary information S1A and S1B**. The same procedures, SP followed by XP were applied for molecular docking of DprE1 inhibitors and X-score re-scoring method was applied to measure the binding affinity of molecules with DprE1 [31]. Glide docking scores and binding affinity (X-score) of top 10 hit-molecules and DprE1 inhibitors are summarized in **Table 1 and 2**, respectively.

The DprE1 inhibitor, **CT319** shows highest docking score $-5.48 \text{ kcal mol}^{-1}$, whereas, **BTZ-N3** shows the lowest docking score $-1.82 \text{ kcal mol}^{-1}$. However, the X-score results show highest binding affinity of **BTz043** ($-9.62 \text{ kcal mol}^{-1}$) with DprE1, whereas, the lowest binding affinity found with **TBA-7371** ($-7.87 \text{ kcal mol}^{-1}$). Among the compounds taken from ChEMBL database, compound ChEMBL2323138 (**C1**) shows highest docking score of ($-10.198 \text{ kcal mol}^{-1}$) with DprE1 and the minimum docking score $-8.795 \text{ kcal mol}^{-1}$ is obtained for ChEMBL1607606 (**C10**). Whereas, the X-score results show highest binding affinity $-10.74 \text{ kcal mol}^{-1}$ with ChEMBL2338605 (**C6**) and lowest affinity $-8.60 \text{ kcal mol}^{-1}$ for ChEMBL1607606 (**C10**).

3.2 ADME property analysis against DprE1 receptor

Another filtering method involves the pharmacokinetic properties (ADME) analysis of hit-molecules [35, 40, 41]. Predicting the bioavailability, toxicity and safety of compounds is an important and integral component of drug designing process [13, 35]. We employed QikProp v5.7 available with Schrodinger 2018-3 to analyse the ADME properties of compounds and compared with DprE1 inhibitors. Results show that compounds, **C5**, **C6**, **C8** and **C10** having the CNS activity with the potential range of drug

molecules -1 to 0 (**Table 3**). All hit-molecules having molecular weight <500. The optimal range value recommended for the lipophilicity (QPlogPo/w) of compound is between -2.0 – 6.5. Result shows that all 10 molecules having QPlogPo/w <6.5, however, the lowest value (0.199) is observed for **C4**. Whereas, the higher QPlogPo/w value 5.108 is obtained for **C5**. The QPlogS (potential range –6.5 – 0.5) defines the aqueous solubility of compounds which are observed within the favourable range for all 10 compounds. The recommended range for predicting IC50 value for blockage of hERG K⁺ channel is QPlogHERG <-5 which is well satisfied by all compounds. The compounds having the predicted apparent Caco-2 cell permeability test (QPlogCaco) > 500 is recommended. Out of 10 molecules, only four compounds, **C5**, **C6**, **C8** and **C10** having a value range > 500. The recommended range for QPlogBB is -3.0 – 1.2 which is observed favourable for all molecules. Similarly, it is observed that all hit-molecules obeyed the drug likeness properties RO5 (Lipinski's rule of five) and found within the recommended range for human oral absorption (PHOA). The ADME analysis of DprE1 inhibitors shows that all 13 compounds are lying within the recommended ranges for predicted ADME descriptors (**Table 4**).

3.3 In silico drug-likeness and toxicity predictions

The physiochemical properties, toxicity, tumorigenicity and mutagenesis risk of the compounds are investigated by OSIRIS Property Explorer [33] and compared with the DprE1 inhibitors (**Table 5 and 6**). Results show that out of 10 hit-molecules, 9 compounds are estimated as no risk for mutagenicity (MUT and tumorigenicity (TUMO), whereas, **C2** shows higher risk for both MUT and TUMO test (**Table 5**). All 10 compounds show no risk for irritation (IR), however, reproductive development (REP) toxicity result shows high risk for **C6**, whereas, **C9** shows medium risk for REP. The drug score (DS) is representing the combined score value of compounds solubility, polar surface area, toxicity, drug-likeness and CLogP which define the overall sensitivity of drug molecules. Result shows highest DS

value 0.79 for the hit-molecule, **C4** and the least observed DS score (0.14) for **C2**. The four compounds, **C5**, **C6**, **C8** and **C10** which successfully cross the ADME test, show moderate DS score ranges 0.26 – 0.42. The compound **C8** shows higher DS score 0.42, **C5**, **C8** and **C10** predicted as no risk for toxicity parameters. Whereas, **C6** predicted as high toxic risk for reproductive effect (REP), as the chemical scaffold contains ketone moiety. **Table 6** shows the toxicity and drug-likeness parameters index of DprE1 inhibitors. The inhibitors, Ty38c and 4AQs predicted as higher risk for mutagenesis, whereas, VI-9376 shows medium risk and other inhibitors observed as no risk. All selected inhibitors having no risk for TUMO. Only, two inhibitors show medium risk for IR and all are predicted as no risk for REP. The predicted DS score of inhibitors ranges 0.22 – 0.88. The inhibitor CT319, co-crystallized with DprE1 X-ray structure shows moderate range of DS score 0.37, however, no risk is observed for toxicity parameters.

3.4. Molecular Interactions

The crystal structure of DprE1 consists of well separated, conversed FAD-domain and the substrate binding domain. The deeply buried FAD-domain is composed of an α/β fold, formed by the residues belonging to N-terminus (residues 7-196) and C-terminus (413-461). The substrate binding domain is extended from flavin at centre to surface, orchestrated by anti-parallel β -sheets (β 10-16) and helices (α 5, 9 and 10). The wide-open active site of DprE1 is governed by two loops at surface which facilitate the accessibility and flexibility of ligand binding. We find that the top 10 hit-molecules (**C1-C10**) taken from ChEMBL shared a common interaction with active site residues which is summarized in **Table 1**. The 2D molecular interaction of hit-molecules C1-C10 and inhibitors at the active site of DprE1 are shown in **Supplementary Figure S2 and S3**. The active site amino acid residues involved in interactions with DprE1 inhibitors are tabulated in **Table 2**.

The co-crystallized structure of DprE1-CT319 shows that the inhibitor at the active site is largely stabilized by the hydrophobic interactions. Trifluoromethyl moiety of **CT319** involved in hydrophobic interaction with Lys134 and Tyr314 and the nitro-benzene is stabilized with Lys317 and Val365. The nitro (NO₂)-group shows H-bond with His312 and Lys418, and the phenylethyl moiety forms hydrophobic interaction with Tyr60, Trp230, Phe320, Leu363, Val365 and FAD present in the vicinity of catalytic domain. Furthermore, the structural studies of BTZs derivative inhibitors demonstrated that the hydrophobic amino acid residues at the active site, Trp60, Gly117, His132, Gly133, Lys134, Ser228, Phe231, Tyr314, Leu317, Phe320, Gln321, Trp323, Asn324, Gln334, Q336, Leu363, Val365, Phe366, Lys367, Phe369, Asn385, Ile386, Cys387, Asp389 and Lys418 are critical for the ligand recognition. And, some inhibitors are covalently linked with C387. Molecular docking result shows that at the active site of DprE1, CT319 forms H-bond with Gln336, Asn385, Lys418, alkyl interactions with Tyr314, π -alkyl interaction with Leu317, Leu363, Cys387 and the π -Sigma bond with Val365 (**Figure 2**). The trifluoromethyl moiety of CT319 interacting with Pro116, Gly133, Lys134 and Tyr314, whereas, the both benzene rings are stabilized with hydrophobic and van der Waals interaction interacted with residues: Tyr60, Gly117, His132, Lys134, Ser228, Phe320, Gly321, Lys367, Asp389 which is observed consistent with the co-crystal structure [8, 11].

The top ranked docking score compound **C6** (7-[3-(2-chloroanilino)-2-hydroxy-propoxyl]spiro[chromane-2,1'-cyclopentane]-4-one) shows that the active site residues Tyr60, Cys387, Lys418, Asn385 are involve in H-bonding which interplay a critical role to stabilize the ligand. Whereas, hydrophobic residue, Val365 formed π -Sigma interaction and, Lys134, Tyr314, Leu317 and Leu363 are involved in π -alkyl interactions. And, amino acid residues Leu115, Pro116, Gly133, Ser228, Phe320, Gly321, Trp323, Asn324, Gln336, Lys367, Phe369, Asp389 are effectively engaged in van der Waals interactions with **C6** (**Figure 3**).

The conventional hydrogen bonding interactions are observed consistent with **C5** (1-[(4-benzyloxyphenyl)-cyclopropyl-methyl] imidazole) showing H-bond interaction with Asn324, carbon-hydrogen bond with Tyr60 and alkyl interaction with Leu363, Trp230. The residues Leu317, Lys367 show π -alkyl interaction, Cys387 engaged in π -sulfur interaction and Lys418 and Val365 are involved in π -cation and π -Sigma interactions, respectively. And, the hydrophobic and van der Waals interactions of Gly117, His132, Gly133, Lys134, Ser228, Phe320, Gly321, Gln334, Gln336, Phe369, Asn385, Asp389 provided the additional stability to **C5** at the active site of DprE1 (**Supplementary Figure 4**). The molecular binding of **C8** (1-cyclohexyl-5-oxo-N-(3-phenylphenyl)pyrrolidine-3-carboxamide) at the active site of shows H-bond interaction with Lys418, π -donor H-bonding with Tyr60, the alkyl and π -alkyl interactions with Lys134, Lys367 and Leu317, Leu363, Val365, respectively. And, the hydrophobic and van der Waals interactions with Pro116, Gly117, His132, Gly133, Ser228, Tyr314, Phe231, Phe320, Gly321, Trp323, Asn324, Gln334, Gln336, Phe366, Phe369, Asn385, Ile386, Asp389 at the active of DprE1 (**Supplementary Figure 5**).

The spatial orientation of **C10** (3-(3-hydroxypropyl)-7-(2-thiophen-3-ylethynyl)isochromen-1-one) shows H-bond interaction with Asn324, Cys387, the carbon hydrogen bond with Phe230 and π -alkyl interaction with Leu317, Leu363, Lys367. The thiophen moiety of C10 involve in π -sulfur interactions with His132, Phe369, whereas, Val365 and Lys418 show π -sigma interaction and π -cation interaction. And, the ligand is stabilized by the hydrophobic and van der Waals interactions of residues: Tyr60, Trp66, Gly117, Gly133, Lys134, Gly321, Glu322, Trp323, Arg325, Gln336, Asn385, Val388, Asn389 which is shown in **Supplementary Figure 6**. These molecular docking results suggested that apart from the conventional H-bonding at the active site of DprE1, several other interactions, π -alkyl (Leu317, Lys367), π -sulfur (Cys387), π -Sigma and the van der Waals interactions of amino residues His132, Gly133, Lys134 and Asn385 are critically involved in interactions with lead-molecules which observed consist with the co-crystalized structure of DprE1-**CT319**. Thus, the molecular docking results, binding

affinity scores and pharmacokinetic analysis of hit-molecules suggested that compounds, **C5** (docking score -9.248 kcal/mol, X-score -9.66 kcal/mol), **C6** (-9.211 kcal/mol, X-score -10.74 kcal/mol), **C8** (-9.106 kcal/mol, X-score -10.64 kcal/mol) and **C10** (docking score -8.795 kcal/mol, X-score -8.60 kcal/mol) may be explored as promising candidates for further lead optimization as DprE1 inhibitors.

3.5 Conformational dynamics and stability of protein-ligand complex

The solvent environment around the protein influences the molecular interaction. Thus, the various interactions observed during the molecular docking may or may not exist during the simulation [15, 42]. To examine the conformational stability, dynamics and structural integrity of DprE1 complex with novel hit-molecules, multiple MD simulations were performed in aqueous environment for the period of 100 ns, at 300 K. The conformational dynamics of DprE1 and DprE1-CT319 during the MD simulation used as a control to elucidate the structural stability of DprE1 complexed with ChEMBL compounds. Trajectories obtained from the simulation were further used for the binding free energy estimation of molecules, using MM/GBSA, MM/PBSA and MM/3DRISM [15, 17, 19, 43].

To determine the conformational stability of DprE1 with hit-molecules, we measured all atom C α -RMSD of protein-ligand complexes and compared the results in reference of DprE1-CT319 complex (**Figure 4**). Results show that the structure of DprE1 remains stable with an average change in RMSD value 3.57 ± 0.49 Å. The trajectory of DprE1 archives equilibrium at ~25 ns and a continuous stable equilibrium can be seen up to 100 ns of simulation. RMSD plot of DprE1 complex with inhibitor **CT319** shows that trajectory achieves equilibrium at ~15 ns and the complex structure remains stable for the remaining period of simulation with change in RMSD 2.80 ± 0.26 Å. The RMSD plot of DprE1-**C5** shows a continuous increase in trajectory during initial 0-35 ns. The complex structure remains stable for the period of 35-55 ns and a small drift of 0.5 Å is observed at 60 ns. The RMSD trajectory during 60-100 ns suggested a stabilized structure of DprE1-**C5** complex for the last 40 ns of simulation. The trajectory

of DprE1-C6 achieves equilibrium in 0-15 ns and remains stable till the simulation finished at 100 ns. We observed a consistent and overlapped RMSD plot of DprE1-C6 with DprE1-CT319 complex. The RMSD plot of DprE1-C8 shows initial perturbation in structure during 0-25 ns, however, reaches to equilibrium at ~30 ns, after that the conformational dynamics remains stable around RMSD 3.66 ± 0.37 Å. Although the structure of DprE1-C10 achieves equilibrium earlier at ~10 ns and remains stable up to 70 ns, however, we observed structural adjustment with the drift of ~1 Å at 75 ns and the structural dynamics remains stable with an average change of RMSD value 3.18 ± 0.47 Å.

To understand the spatial stability of ligand molecules at active site of DprE1, we also calculated the time evolution plot of distance of hit-molecules and inhibitor CT319 from the centre of binding pocket, as shown in **Figure 5**. We observed that the average distance of CT319, C6 and C8 remains quite stable suggesting that ligand is spatially well occupied at active site and stabilized with molecular interaction, during the simulation. The compound C5 shows continuous drop down in distance during 0-40 ns and it stabilized at distance ~4 Å which is seen up to 100 ns. The distance plot of C10 shows fluctuating behaviour which suggested the spatial adjustment at the binding pocket, thus, we observed a small drift of 1 Å in RMSD plot of DprE1-C10 complex.

The conformational order parameter, radius of gyration (R_g) represents structural compactness and integrity of a protein structure [44]. The R_g plot shows that all five complexes are stabilized around average R_g value ~21-23 Å, suggesting that all the ligand molecules were well occupied at the binding pocket of DprE1 during the simulation (**Figure 6**). Similar to RMSD results, we observed slightly higher R_g value 22.02 ± 0.13 Å for the ligand unbound structure of DprE1. The complex of DprE1-C6 shows lowest R_g value 21.75 ± 0.06 Å and highest 22.28 ± 0.17 Å for DprE1-C5, whereas, the structure of DprE1-CT319 is stabilized around R_g score 22.10 ± 0.07 Å. The marginal differences in R_g value of DprE, ligand bound, and unbound structures suggested the stable interaction of novel hits molecules at the active site of protein. These results provide a clear evidence that the selected ligands are well accommodated in

binding pocket, having consistent interactions with active site residues as observed during the molecular docking.

We further investigated the conformational fluctuations and local dynamics through the calculation of average fluctuation of each amino acid residue of DprE1. The RMSF plot of all C^α-atoms of DprE1 and docked complexes with ligands are shown in **Figure 7**. In this **Figure 7**, we can see the comparative results of DprE1, and each ligand bind complex with DprE1. Results show that the average fluctuation of residues are reduced on binding of ligands at the binding pocket of DprE1 which is suggesting a favourable molecular interaction. We observed that average fluctuation of C-terminal residues is increased on binding of **CT319** and **C5**. The RMSF plots show higher mobility for residues 150-200, belonging to β_{8-10} , which can be seen in **CT319**, **C5** and **C6**. Whereas, DprE1-**C8** and DprE1-**C10** show lower average fluctuation in compared to FAD bound DprE1. However, the secondary structure analysis plots using DSSP suggested that no significant conformational changes are observed in the secondary structure, upon the binding of ligands during simulation which provide an elegance evidence of stable molecular interactions of ligands with DprE1 (**Supplementary Figure 7**).

3.6 Hydrogen bond analysis

The structure of protein is largely stabilized with the network of H-bond which plays critical role in the conformational adaptability, mobility and interaction with biomolecules. Apart from the structural stability, H-bond interactions are crucial in molecular recognition and protein-ligand interactions. Thus, we analysed H-bond interactions between protein and ligands, using cpptraj module of Amber with distance cut off 3.5 Å and angle cut off 135°. Results show that average two to four H-bond interactions are involved in DprE1 interactions with ligands (**Figure 8**). However, DprE1-**C6** shows highest six H-bond interactions, on an average four H-bond interactions are involved with **CT319** and the least H-bond interactions are observed with **C8**. For the reference inhibitor CT319, we find H-bond interactions

between one O atom and H atoms at zeta position of Lys418 remains stable throughout the simulation (**Figure 8A**). The H-bonds between two O atoms and H atoms at delta position of Asn385 also remain intact and N atom shows H-bond with H atom at delta position of Asn385, however, the H-bond observed in molecular docking with Gln336 and O atoms lost during the MD simulation. The compound **C6** shows H-bond between H atoms at zeta position of Lys418 and O atom observed intact during the simulation, whereas, the double bonded O atom which is involved in H-bond interaction with H atom at delta position of Asn385 is observed to form additional H-bonds with H atoms at epsilon position of Gln336 and His132 (**Figure 8C**). In **C5**, the H-bond formed between N atom and Asn324 is broken and new H-bonds are formed between this N-atom and H atoms at zeta position of Lys418 during the simulation (**Figure 8B**). The O atom present in **C5** was also found to form H-bonds with H atoms at epsilon position of His132, Gln336 and at delta position of Asn385. During simulation we find that the H-bond between O atom lying at the end of the alkyl chain of **C10** and H atom at delta position of Asn324 remains consistent during the simulation (**Figure 8E**), However, H-bond formed between double bonded O atom (**connected to O containing ring**) and Cys387 is lost during the simulation but this double bonded O atom forms new H-bonds with H atom at zeta position of Lys418 and H atom at epsilon position of His132. **Figure 8D** shows that the H-bond between O atom (present between two rings) of compound **C8** and H atoms present at zeta position of Lys418 remains stable during the simulation. While the simulation, we observed the formation of an additional H-bond with H atom at delta position of Asn385.

3.7 Essential Dynamics

We further performed essential dynamics (ED) analysis to understand the dynamics of protein-ligand complexes. ED analysis involves representation of collective motion of the most variable region of protein in terms of two principal components PC1 and PC2. The projection of each protein-ligand complex trajectory along with native protein onto two principal components PC1 and PC2 is shown in **Figure 9**. From figure 9, we can see that two features are very apparent. Firstly, the protein, DprE1 alone

explores a wide range of conformations in water. Secondly, except hit-molecules, **C5** and **C8**, DprE1 complexed with **CT319**, **C6** and **C10** covers a smaller region on the plot particularly along PC1 plane. These observations thus support the idea of decrease in flexibility in the presence of DprE1 inhibitor, **CT319** and compounds, **C6** and **C10**. In the presence of CT319, DprE1 is restricted to small excursions slightly away from its initial conformation. The finding of a strong restriction in the size of the explored conformational space with only a minor reduction in RMSF which can be seen in **Figure 7**, indicates that local fluctuations take place but that collective motions have been compromised or more likely slowed down in the presence of **CT319**, **C6** and **C10** (**Figure 9A, 9C and 9E**) as compared to DprE1 complex with **C5** and **C8** (**Figure 9B and 9D**). Thus, the ED results along with H-bond interactions suggested that protein-ligands interactions remain consistent during the simulation, however, the most stable conformational dynamics is observed for CT319, C6 and C10.

3.8 Binding free energy analysis

The quantitative assessment of molecular binding interaction of DprE1 inhibitor **CT319** and lead molecules **C5**, **C6**, **C8** and **C10** are estimated using three different methods for the molecular theory of solvation, MM/GBSA, MM/PBSA and MM/3DRISM-KH. MM/GBSA and MM/PBSA calculations are performed on the 5000 frames taken from the last 50 ns of the MD simulation [12, 18, 43]. Considering the large computational cost, 100 equally spaced frames taken from the last 50 ns of simulation are used for MM/3DRISM-KH analysis. The result of these three calculations are given in **Table 7, 8 and 9**, respectively. All three methods show that van der Waals ($\Delta\text{EMM-VDW}$) and electrostatic ($\Delta\text{EMM-EEL}$) components of molecular mechanics (MM) force field energy always favour the protein-ligand binding for all five compounds. And, the polar part of solvation free energy ($\Delta\text{GSOL-GB}/\Delta\text{GSOL-PB}/\Delta\text{GSOL-POL}$) estimated by all three methods suggested the lesser contribution for dynamic stability of DprE1 interactions with selected compounds and inhibitor. It is noted that $\Delta\text{EMM-EEL}$ values estimated by MM/PBSA method are nearly four times smaller as compared to corresponding MM/GBSA and

MM/3DRISM-KH values because MM/PBSA calculations were done with solute dielectric constant 4. Furthermore, MM/3DRISM method also gives the energetic and entropic component of the solvation free energy (**Table 9**) which shows that it is the entropic part of solvation free energy not energetic, which favours the protein-ligand binding for all the ligands. Thus, the all three methods do not give the same order of binding free energy for the protein-ligands, but the comparison of predicted binding free energies provide an important clue to evaluate the relative stabilities and flexibilities of compounds at the active site of DprE1. In the **table 7**, results of MM-GBSA analyses show the higher value of combined ΔG (-41.28 ± 3.51 kcal mol⁻¹) for compound **C6**, whereas, lowest estimated ΔG for CT319 (-35.31 ± 3.44 kcal mol⁻¹). The compound **C5** and **C10** shared the almost similar binding free energies (ΔG ~ 36 kcal mol⁻¹) and the slightly better binding energy predicted for **C8** (-40.75 ± 3.86 kcal mol⁻¹). Moreover, the binding energies estimated by MM-PBSA (**Table 8**) also suggested the major contribution of non-polar solvation energies in the molecular interaction of compounds at the active site of DprE1. We find that the ΔG values for lead compounds ranges $-16.08 - -22.36$ kcal mol⁻¹, whereas, the estimated ΔG -13.67 ± 2.65 kcal mol⁻¹ for CT319. In another analysis, the molecular theory solvation, MM-RISM-KH which yields the broader picture of molecular interactions on solvation structure and implication of thermodynamics from the first principles, accounts for solvent and biomolecules to describe the relative binding affinities [17]. **Table 9** shows that binding interaction of compounds stayed stable as perceived from the molecular docking, however, results again reveal the highest ΔG value -10.33 ± 5.70 kcal mol⁻¹ for **C6** and lowest for **CT319** (ΔG value -3.40 ± 4.06 kcal mol⁻¹). Furthermore, the binding free energy approximation by all three methods suggested the larger contribution of van der Waals energies for ligands interactions and stability at the active site of DprE1. Collectively, the results demonstrated that all four hit-molecules C5, C6, C8 and C10 have better binding affinity with DprE1 as compared to inhibitor CT319 (**Figure 10**). Thus, the lead optimization of selected four compounds from ChEMBL

chemical database may provide a new endeavour for the development of DprE1 inhibitors in MTB therapy.

4. Conclusion

In conclusion, we have explored structure based virtual screening for the identification of promising chemical entities as DprE1 inhibitors from ChEMBL database. Initial sorting of compounds results in the selection of 30,789 small molecules which are suggested for the anti-mycobacterial activity. The three steps molecular docking and binding affinity estimation processes lead to the selection of bioactive 10 hit-molecules. Similar procedures were applied on the selected 13 DprE1-inhibitors which were used to compare the results with hit-molecules. The extensive evaluation of pharmacokinetic profile and drug-likeness properties analyses using ADME, toxicity and OSIRIS properties explorer suggested that four chemical entities, **C5 (ChEMBL2441313)**, **C6 (ChEMBL2338605)**, **C8 (ChEMBL441373)** and **C10 (ChEMBL1607606)** may be explored as potential candidates for the lead optimization as DprE1 inhibitors. To determine the conformational stability of hit-molecules at the active site of DprE1 in aqueous environment, multiple MD simulation were performed on the complex of DprE1 with lead molecules and inhibitor **CT319**. The binding free energy estimation using MM/PBSA, MM/GBSA and 3D-RISM-KH revealed that compounds **C5**, **C6**, **C8** and **C10** show better binding affinity as compared to DprE1 inhibitors. Thus, our comparative studies suggested that the selected compounds (**C5**, **C6**, **C8** and **C10**) could be further investigated as novel lead molecules for the rational drug designing of DprE1-inhibitors in MTB therapy.

Acknowledgements

NK and RK are thankful to UGC and CSIR, respectively, for their fellowships. Authors gratefully acknowledge the computational facility funded by Science and Engineering Research Board (SERB), Government of India (Ref. No.: YSS/2015/000228/LS).

Conflict of Interest: Authors declare there is no conflict of interest.

References:

- [1] Harding, E. WHO global progress report on tuberculosis elimination. *Lancet Respir Med.* 2020, 8, 19.
- [2] Cole, S.T., Riccardi, G. New tuberculosis drugs on the horizon. *Curr Opin Microbiol.* 2011, 14, 570-6.
- [3] Reid, M.J.A., Arinaminpathy, N., Bloom, A., Bloom, B.R., Boehme, C., Chaisson, R., et al. Building a tuberculosis-free world: The Lancet Commission on tuberculosis. *Lancet.* 2019, 393, 1331-84.
- [4] Koul, A., Arnoult, E., Lounis, N., Guillemont, J., Andries, K. The challenge of new drug discovery for tuberculosis. *Nature.* 2011, 469, 483-90.
- [5] Foo, C.S., Lechartier, B., Kolly, G.S., Boy-Rottger, S., Neres, J., Rybniker, J., et al. Characterization of DprE1-Mediated Benzothiazinone Resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother.* 2016, 60, 6451-9.
- [6] Dye, C., Williams, B.G. Criteria for the control of drug-resistant tuberculosis. *Proc Natl Acad Sci U S A.* 2000, 97, 8180-5.
- [7] Alderwick, L.J., Harrison, J., Lloyd, G.S., Birch, H.L. The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. *Cold Spring Harb Perspect Med.* 2015, 5, a021113.
- [8] Chikhale, R.V., Barmade, M.A., Murumkar, P.R., Yadav, M.R. Overview of the Development of DprE1 Inhibitors for Combating the Menace of Tuberculosis. *J Med Chem.* 2018, 61, 8563-93.
- [9] Kolly, G.S., Boldrin, F., Sala, C., Dhar, N., Hartkoorn, R.C., Ventura, M., et al. Assessing the essentiality of the decaprenyl-phospho-d-arabinofuranose pathway in *Mycobacterium tuberculosis* using conditional mutants. *Mol Microbiol.* 2014, 92, 194-211.
- [10] Makarov, V., Lechartier, B., Zhang, M., Neres, J., van der Sar, A.M., Raadsen, S.A., et al. Towards a new combination therapy for tuberculosis with next generation benzothiazinones. *EMBO Mol Med.* 2014, 6, 372-83.
- [11] Batt, S.M., Jabeen, T., Bhowruth, V., Quill, L., Lund, P.A., Eggeling, L., et al. Structural basis of inhibition of *Mycobacterium tuberculosis* DprE1 by benzothiazinone inhibitors. *Proc Natl Acad Sci U S A.* 2012, 109, 11354-9.
- [12] Horoiwa, S., Yokoi, T., Masumoto, S., Minami, S., Ishizuka, C., Kishikawa, H., et al. Structure-based virtual screening for insect ecdysone receptor ligands using MM/PBSA. *Bioorg Med Chem.* 2019, 27, 1065-75.
- [13] Prakash, A., Kumar, K., Islam, A., Hassan, M.I., Ahmad, F. Receptor Chemoprint Derived Pharmacophore Model for Development of CAIX Inhibitors. *Journal of Carcinogenesis & Mutagenesis* 2013, 5, 1-9.
- [14] Lionta, E., Spyrou, G., Vassilatis, D.K., Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem.* 2014, 14, 1923-38.
- [15] Wang, C., Greene, D., Xiao, L., Qi, R., Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front Mol Biosci.* 2017, 4, 87.
- [16] Prakash, A., Dixit, G., Meena, N.K., Singh, R., Vishwakarma, P., Mishra, S., et al. Elucidation of stable intermediates in urea-induced unfolding pathway of human carbonic anhydrase IX. *J Biomol Struct Dyn.* 2018, 36, 2391-406.
- [17] Genheden, S., Luchko, T., Gusarov, S., Kovalenko, A., Ryde, U. An MM/3D-RISM approach for ligand binding affinities. *J Phys Chem B.* 2010, 114, 8505-16.
- [18] Kumari, R., Kumar, R., Lynn, A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model.* 2014, 54, 1951-62.
- [19] Wang, J., Hou, T. Develop and test a solvent accessible surface area-based model in conformational

entropy calculations. *J Chem Inf Model*. 2012, 52, 1199-212.

[20] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al. The Protein Data Bank. *Nucleic Acids Res*. 2000, 28, 235-42.

[21] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004, 25, 1605-12.

[22] Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013, 27, 221-34.

[23] Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012, 40, D1100-7.

[24] Sadowski, J., Gasteiger, J., Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *Journal of Chemical Information and Computer Sciences*. 1994, 34, 1000-8.

[25] Guillemette, G., Poitras, M., Boulay, G. Two Ca²⁺ transport systems are distinguished on the basis of their Mg²⁺ dependency in a post-nuclear particulate fraction of bovine adrenal cortex. *Cell Calcium*. 1991, 12, 51-60.

[26] Shelley, J.C., Cholleti, A., Frye, L.L., Greenwood, J.R., Timlin, M.R., Uchimaya, M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des*. 2007, 21, 681-91.

[27] Jorgensen, W.L., Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A*. 2005, 102, 6665-70.

[28] Jorgensen, W.L., Maxwell, D.S., Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*. 1996, 118, 11225-36.

[29] Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*. 2004, 47, 1750-9.

[30] Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004, 47, 1739-49.

[31] Wang, R., Lai, L., Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des*. 2002, 16, 11-26.

[32] Wallace, A.C., Laskowski, R.A., Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. 1995, 8, 127-34.

[33] Sander, T., Freyss, J., von Korff, M., Reich, J.R., Rufener, C. OSIRIS, an entirely in-house developed drug discovery informatics system. *J Chem Inf Model*. 2009, 49, 232-46.

[34] Joung, I.S., Cheatham, T.E., 3rd. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B*. 2008, 112, 9020-41.

[35] Wang, Y., Xing, J., Xu, Y., Zhou, N., Peng, J., Xiong, Z., et al. In silico ADME/T modelling for rational drug design. *Q Rev Biophys*. 2015, 48, 488-515.

[36] Luthra, P.M., Kumar, R., Prakash, A. Demethoxycurcumin induces Bcl-2 mediated G2/M arrest and apoptosis in human glioma U87 cells. *Biochem Biophys Res Commun*. 2009, 384, 420-5.

[37] Luthra, P.M., Prakash, A., Barodia, S.K., Kumari, R., Mishra, C.B., Kumar, J.B. In silico study of naphtha [1, 2-d] thiazol-2-amine with adenosine A_{2A} receptor and its role in antagonism of haloperidol-induced motor impairments in mice. *Neurosci Lett*. 2009, 463, 215-8.

[38] Mishra, C.B., Kumari, S., Manral, A., Prakash, A., Saini, V., Lynn, A.M., et al. Design, synthesis, in-silico and biological evaluation of novel donepezil derivatives as multi-target-directed ligands for the

treatment of Alzheimer's disease. *Eur J Med Chem.* 2017, 125, 736-50.

[39] Prakash, A., Luthra, P.M. *Insilico study of the A(2A)R-D (2)R kinetics and interfacial contact surface for heteromerization. Amino Acids.* 2012, 43, 1451-64.

[40] Kumar, N., Srivastava, R., Prakash, A., Lynn, A.M. *Structure-based virtual screening, molecular dynamics simulation and MM-PBSA toward identifying the inhibitors for two-component regulatory system protein NarL of Mycobacterium Tuberculosis. J Biomol Struct Dyn.* 2019, 1-15.

[41] Lipinski, C.A. *Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods.* 2000, 44, 235-49.

[42] Spyraakis, F., Benedetti, P., Decherchi, S., Rocchia, W., Cavalli, A., Alcaro, S., et al. *A Pipeline To Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and Fingerprints for Ligand and Proteins. J Chem Inf Model.* 2015, 55, 2256-74.

[43] Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J.Z.H., et al. *End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev.* 2019, 119, 9478-508.

[44] Nygaard, M., Kragelund, B.B., Papaleo, E., Lindorff-Larsen, K. *An Efficient Method for Estimating the Hydrodynamic Radius of Disordered Protein Conformations. Biophys J.* 2017, 113, 550-7.

Figures:

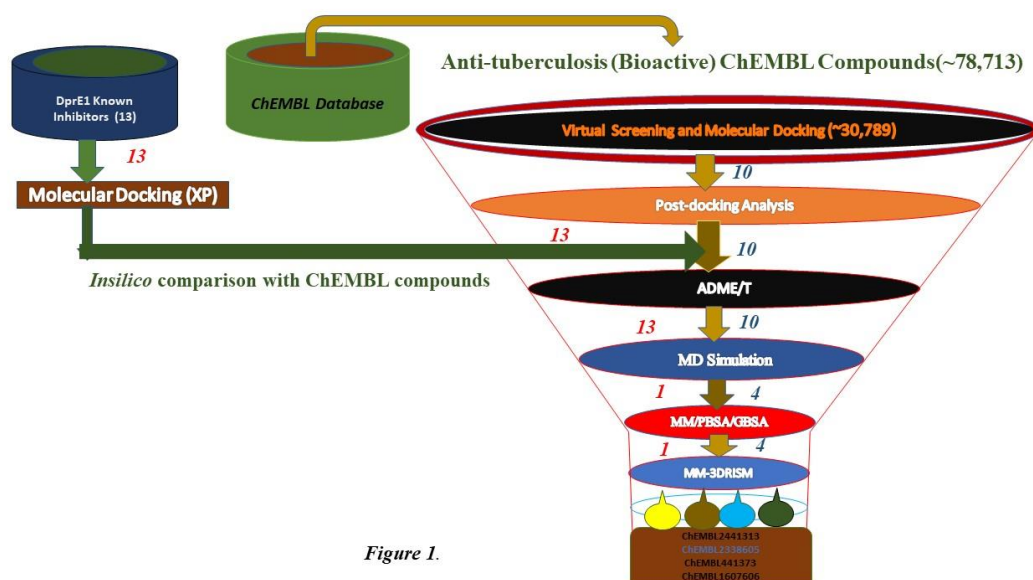


Figure 1.

Figure 1: Flowchart outlined the virtual screening of anti-tuberculosis molecules as DprE1 inhibitor from the chemical compounds database ChEMBL.

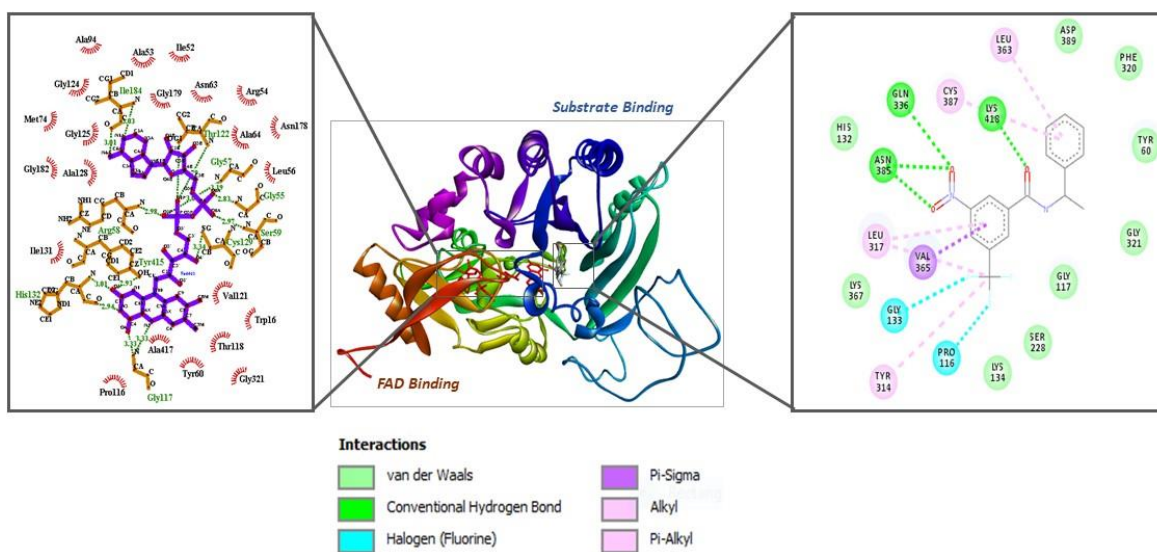


Figure 2.

Figure 2. Docked complex of DprE1 with inhibitor CT319. DprE1 is shown in cartoon and the inhibitor is represented as stick model. The 2D representation of molecular interaction at the active site of DprE1 using Discovery Studio is shown right panel. Left panel shows the active site residues involved in interactions with CT319, obtained from LigPlot.

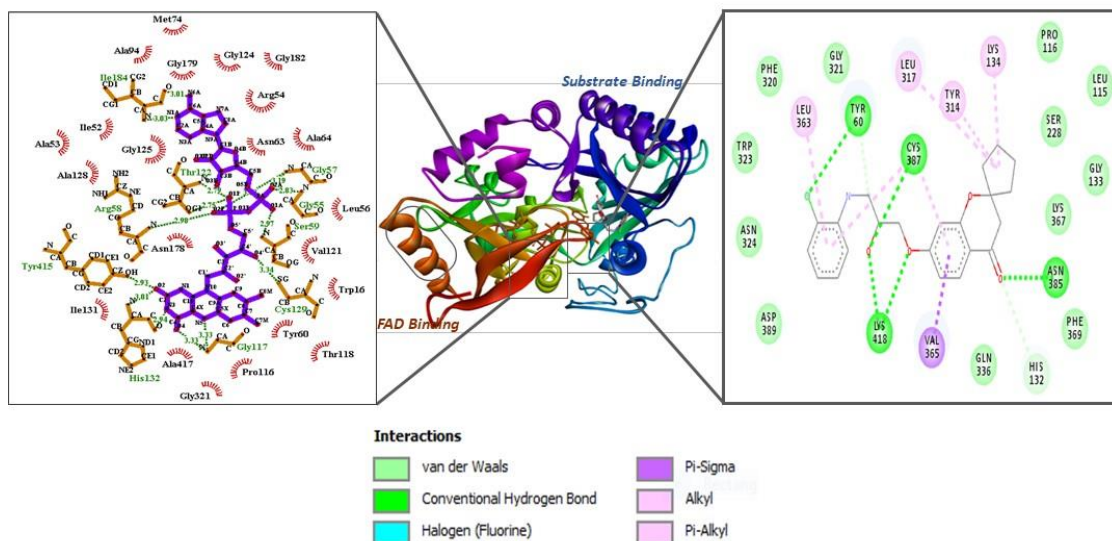


Figure 3.

Figure 3. Cartoon view of DprE1 docked with best hit compound ChEMBL2338605 (C6). Right panel shows the Discovery Studio view of molecular interaction of C6 at the active site of DprE1 and the left panel representing the protein-ligand interactions using LigPlot. The dashed and colored lines representing the different interactions as shown in panel below.

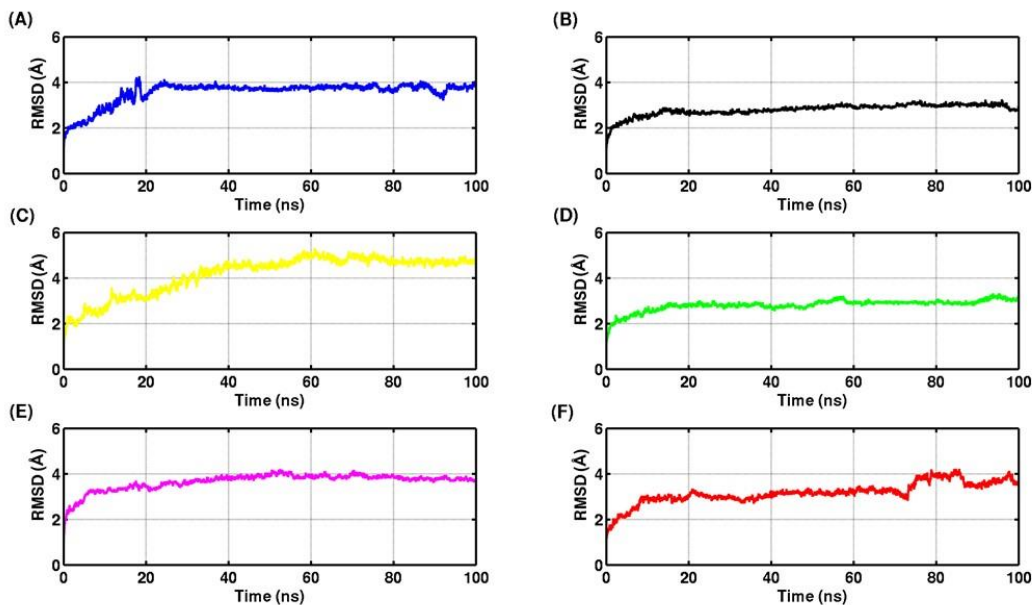


Figure 4.

Figure 4. The time evolution changes in C^α RMSD of (A) DprE1 (blue), and protein-ligand complexes, (B) DprE1-CT319 (black) (C) DprE1-C5 (yellow), (D) DprE1-C6 (green), (E) DprE1-C8 (magenta) and (F) DprE1-C10 (red) in water at 300 K, during the simulation period of 100 ns.

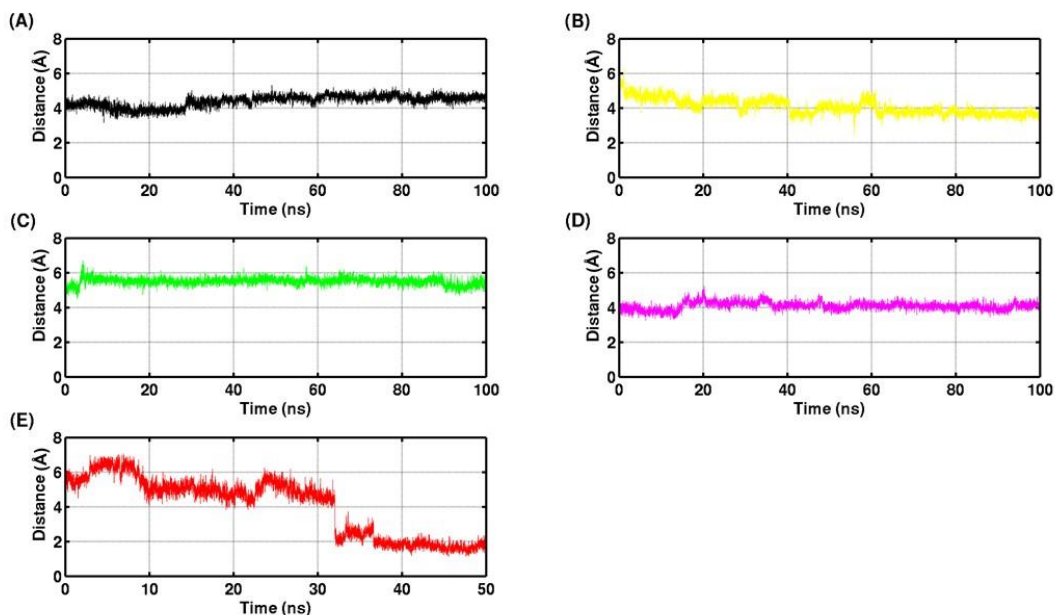


Figure 5.

Figure 5. The time evolution plot of average distance of (A) inhibitor **CT319** (black) and hit-molecules, (B) **C5** (yellow), (C) **C6** (green), (D) **C8** (magenta) and (E) **C10** (red), from the active site center of *DprE1*.

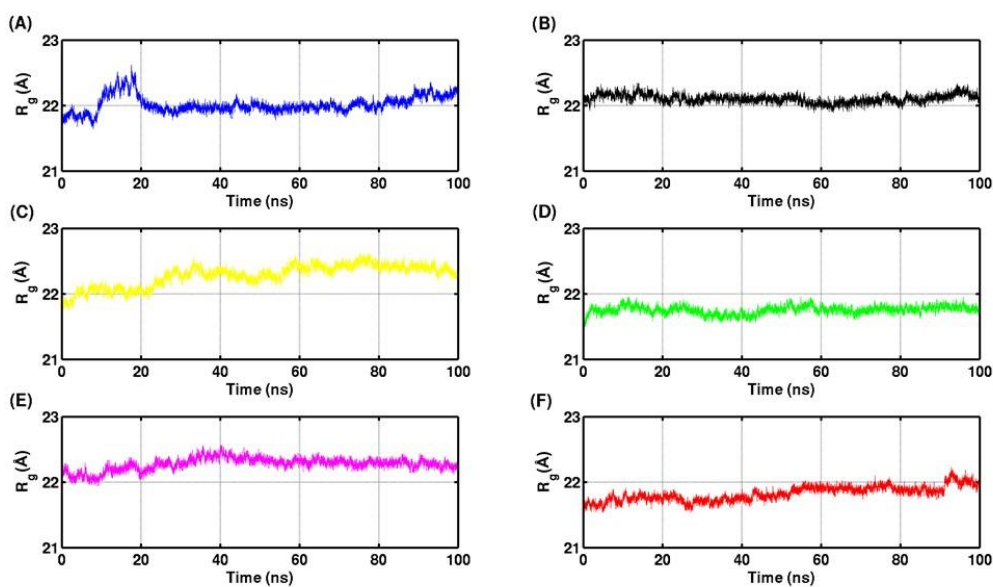


Figure 6.

Figure 6. Radius of gyration (R_g) plot of (A) *DprE1*, and protein-ligand complexes, (B) *DprE1*-**CT319** (C) *DprE1*-**C5**, (D) *DprE1*-**C6**, (E) *DprE1*-**C8** and (F) *DprE1*-**C10**, during the simulation in water at 300 K.

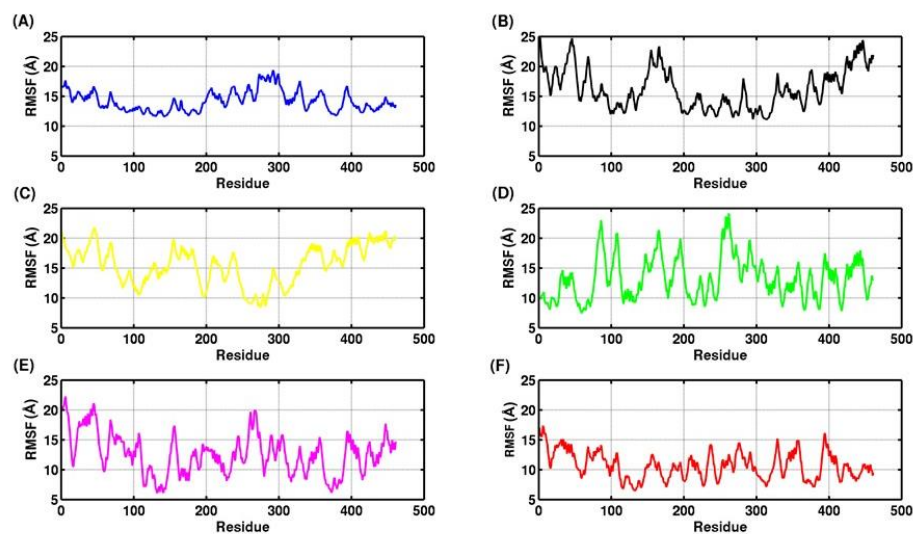


Figure 7.

Figure 7. RMSF (Root mean square fluctuation) plot of all C^α -atoms of (A) DprE1 and the docked complexes with (B) inhibitor CT319 and hit-molecules, (C-F) C5, C6, C8 and C10, respectively.

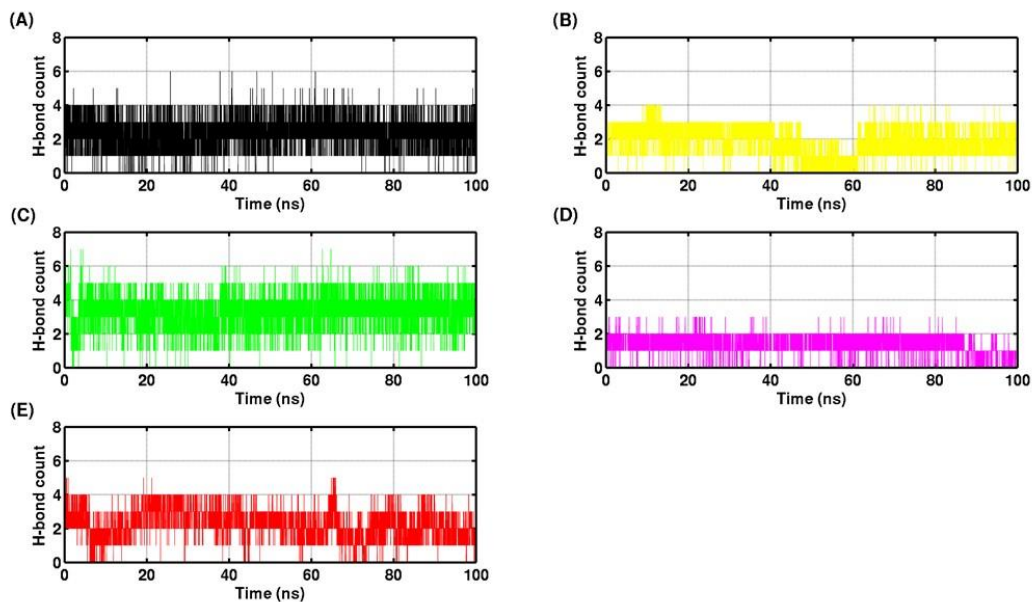


Figure 8.

Figure 8. Hydrogen bond (H-bond) plot representing, the average H-bond counts observed between the DprE1 and ligands, involve in stabilization of molecules at the active site, during the simulation period.

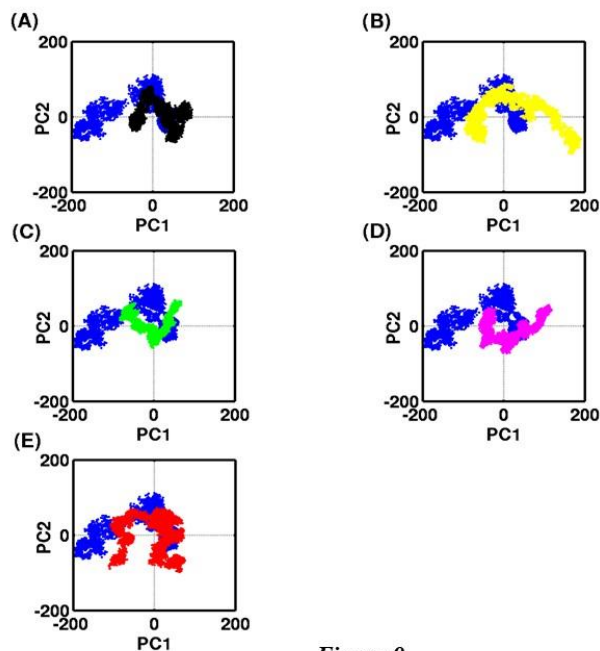


Figure 9.

Figure 9. Principal component analysis (PCA) of DprE1 (blue) and docked complexes with CT319 (black) and hit molecules C5 (yellow), (C) C6 (green), (D) C8 (magenta) and (E) C10 (red). Plots representing the collective motion of protein and docked complexes, using the projections of two principal components PC1 and PC2 calculated from the MD trajectories.

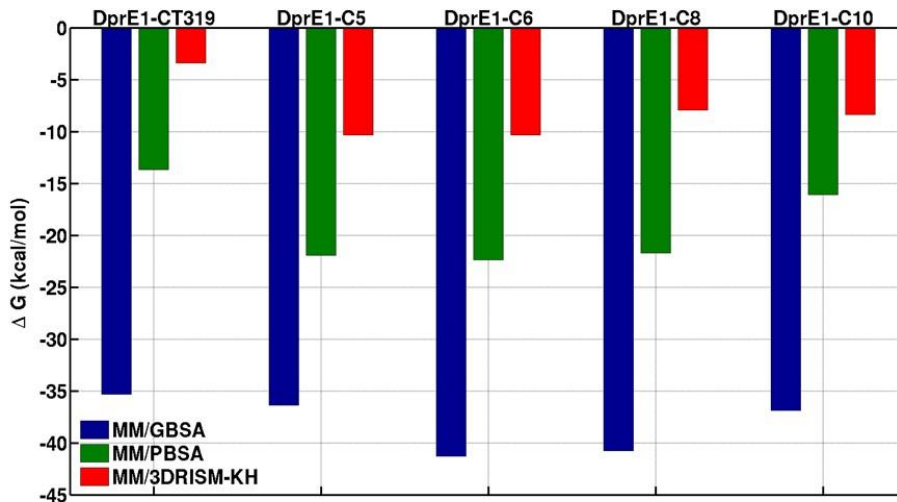
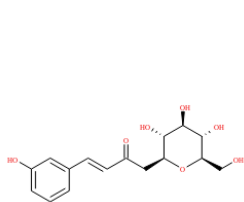


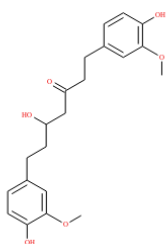
Figure 10.

Figure 10. Binding free energy (ΔG) estimation of the DprE1 complexes with CT319 and hit-molecules (C5, C6, C8 and C10) using theory of solvation, MM/GBSA, MM/PBSA and MM/3DRISM-KH from the MD trajectories.

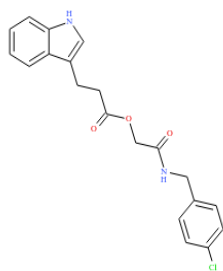
Supplementary Information



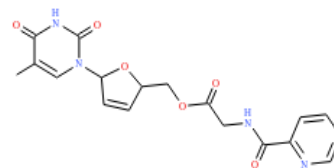
CHEMBL2323138



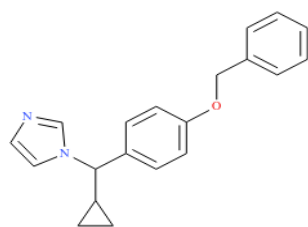
CHEMBL479650



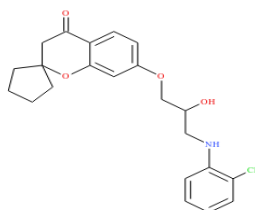
CHEMBL1393315



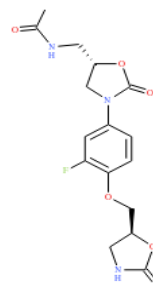
CHEMBL433772



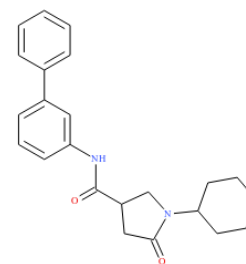
CHEMBL2441313



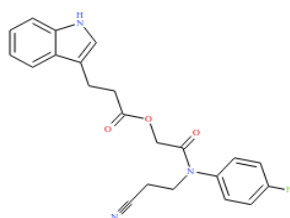
CHEMBL2338605



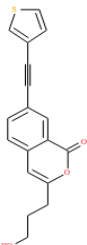
CHEMBL3126094



CHEMBL441373

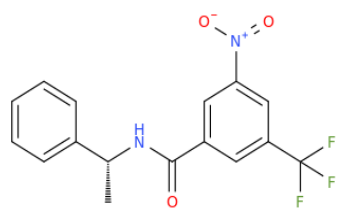


CHEMBL1528618

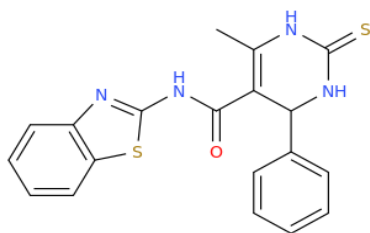


CHEMBL1607606

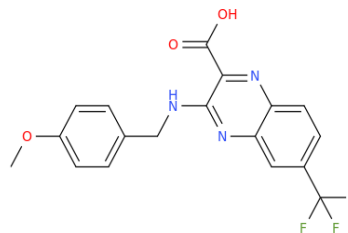
Supplementary Figures 1 (A). 2D-structure of top ten hit-molecules obtained from the ChEMBL database.



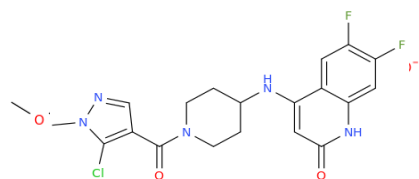
CT319



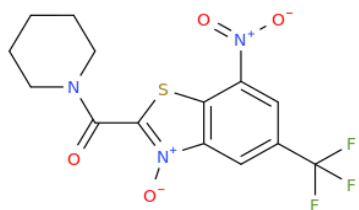
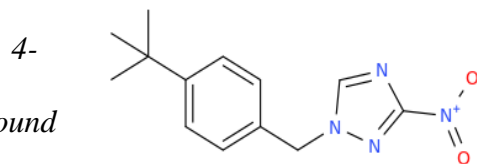
cmp-7a



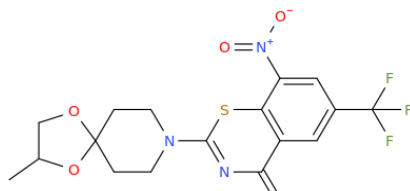
Ty38c



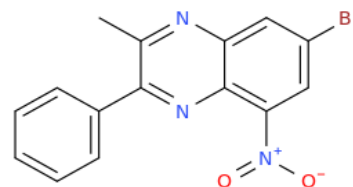
DNBI
Aqs (4-Aminoquinolone
piperidine amides) Compound
377790



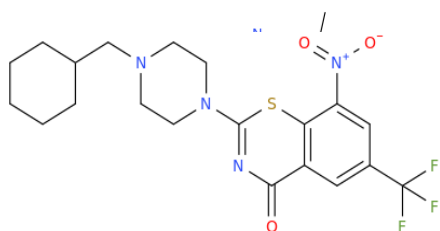
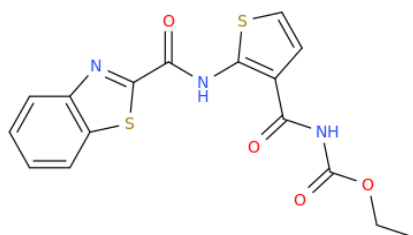
VI-9376



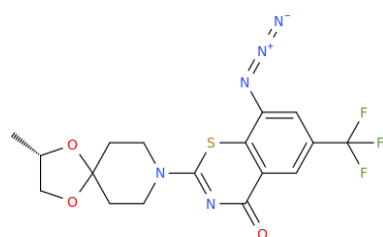
Benzothiazole-N-oxide (BTO)



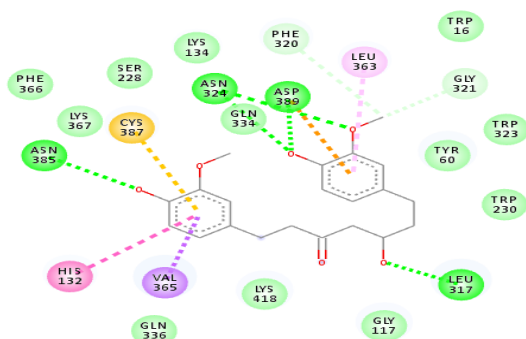
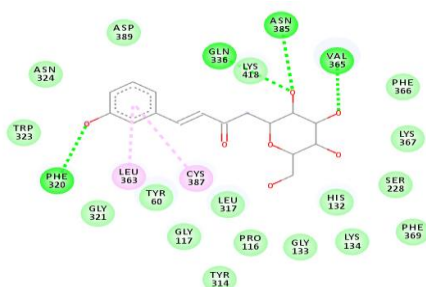
BTZ043



TCA1
TBA-7371
PBTZ169

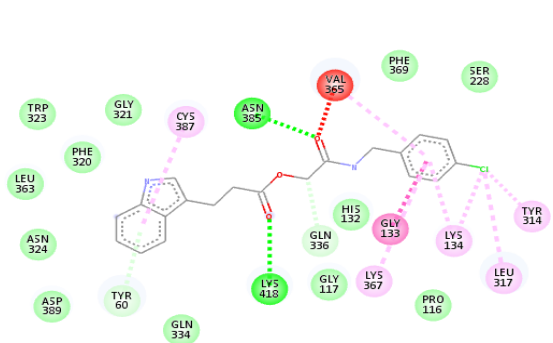


BTZ-N3
Supplementary Figures 1 (B). 2D-structure of DprE1 inhibitors

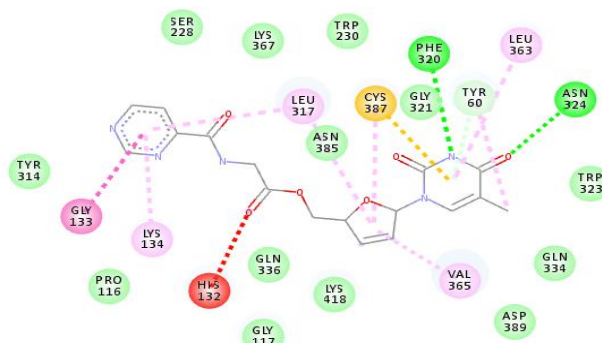


ChEMBL2323138 (C1)

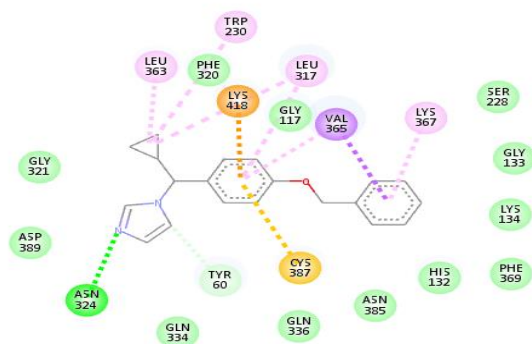
ChEMBL479650 (C2)



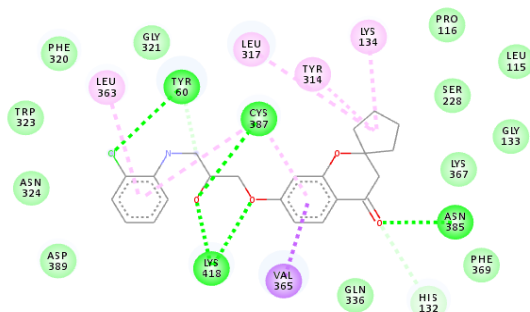
ChEMBL1393315 (C3)



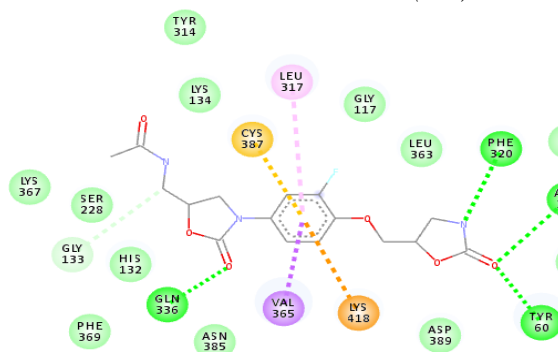
ChEMBL433772 (C4)



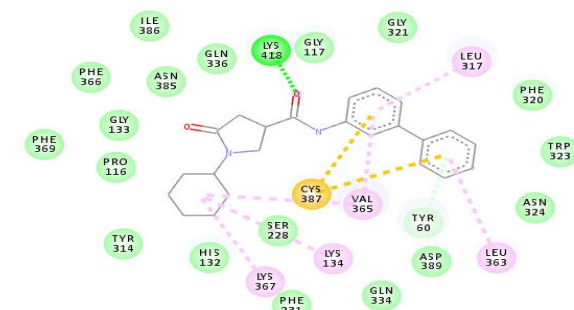
ChEMBL2441313 (C5)



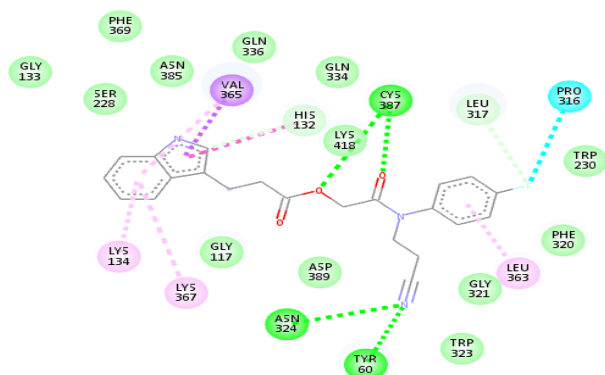
ChEMBL2338605 (C6)



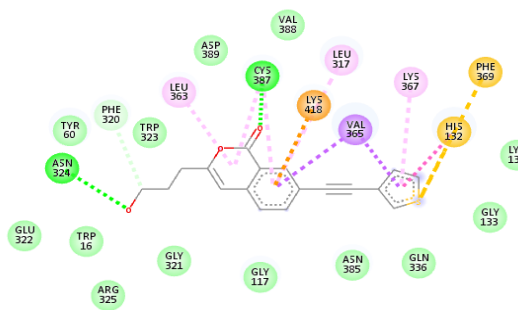
ChEMBL3126094 (C7)



ChEMBL441373 (C8)

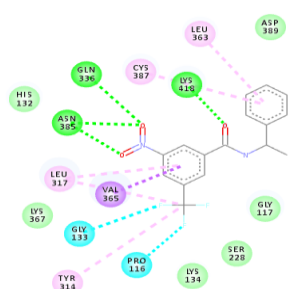


ChEMBL1528618 (C9)

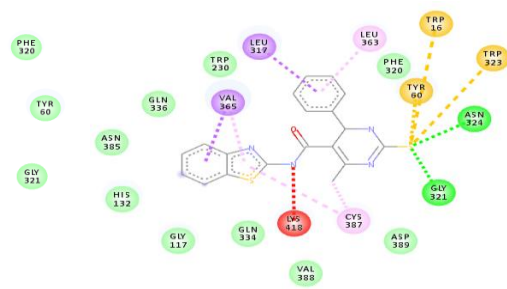


ChEMBL1607606 (C10)

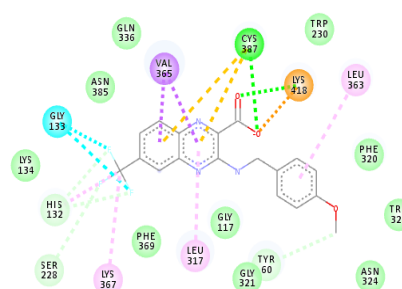
Supplementary Figure S2. 2D molecular interaction representation of the best 10 compounds (C1-C10) at the active site of DprE1.



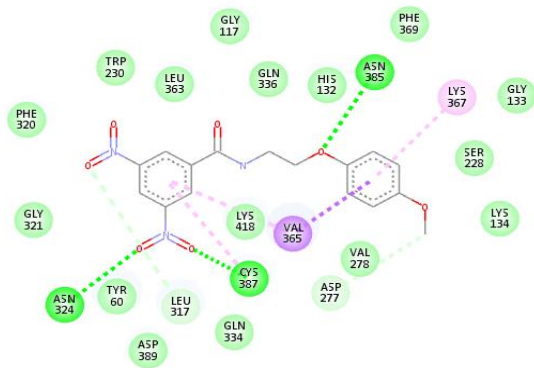
CT319



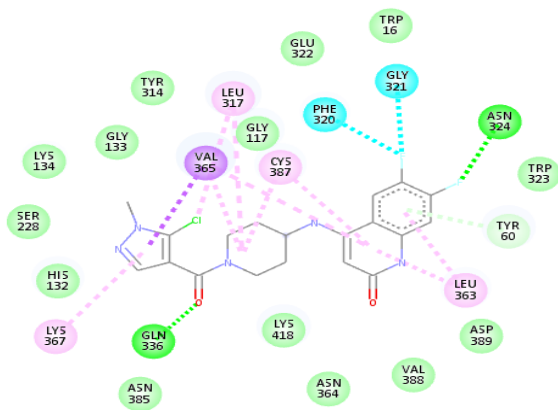
CMP7A



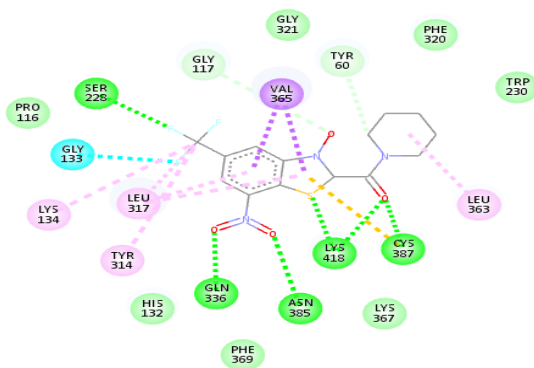
Ty38C



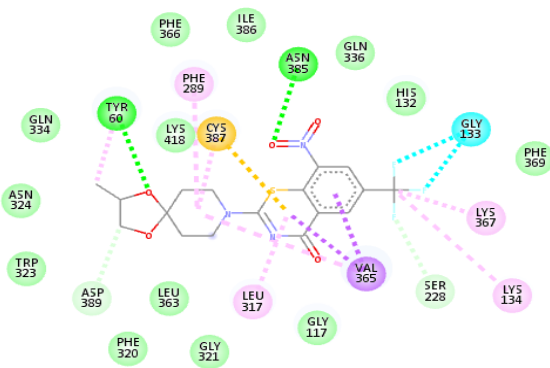
DNBI



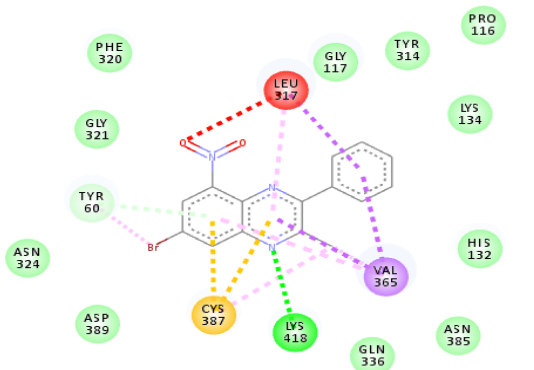
4AQs



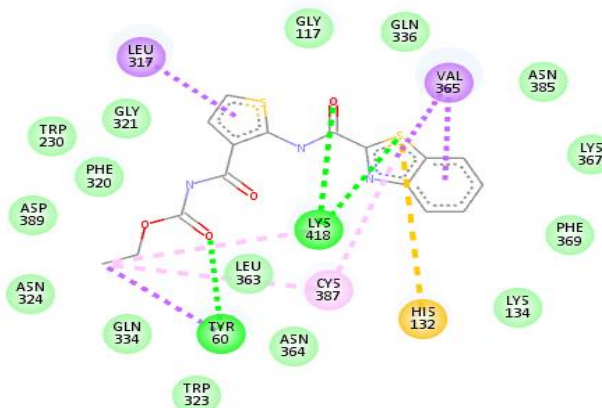
BTO



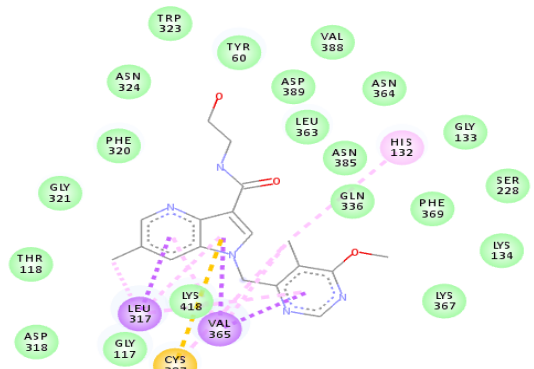
BTZ043



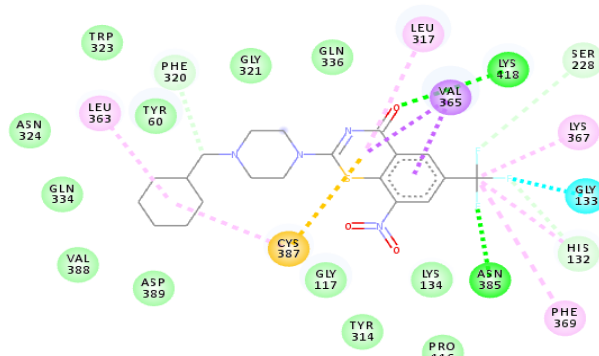
VI-9376



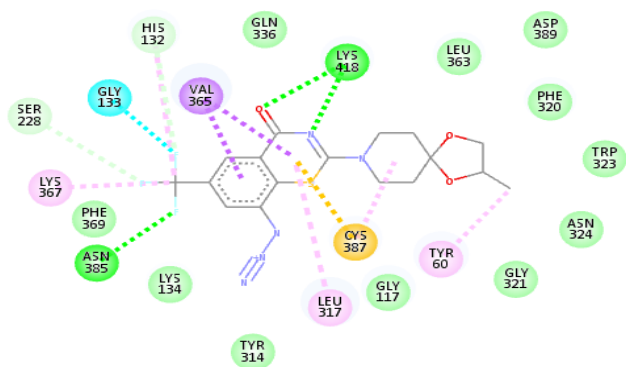
TCAI



TBA-7371

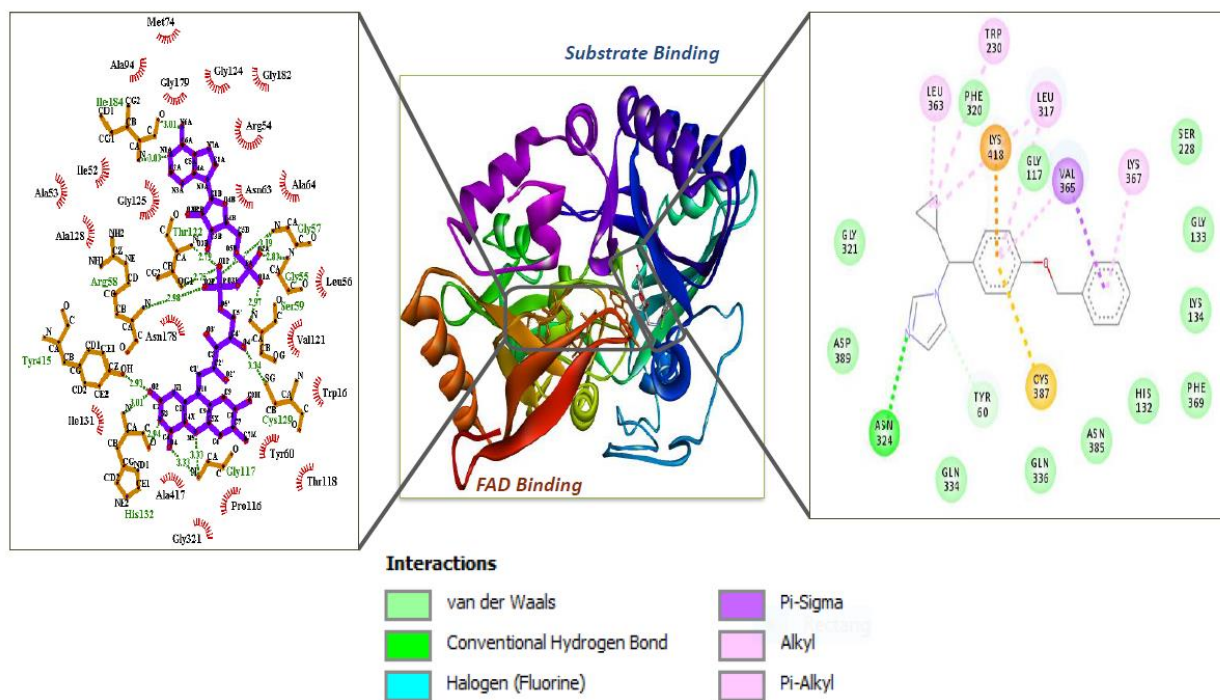


PBTz169

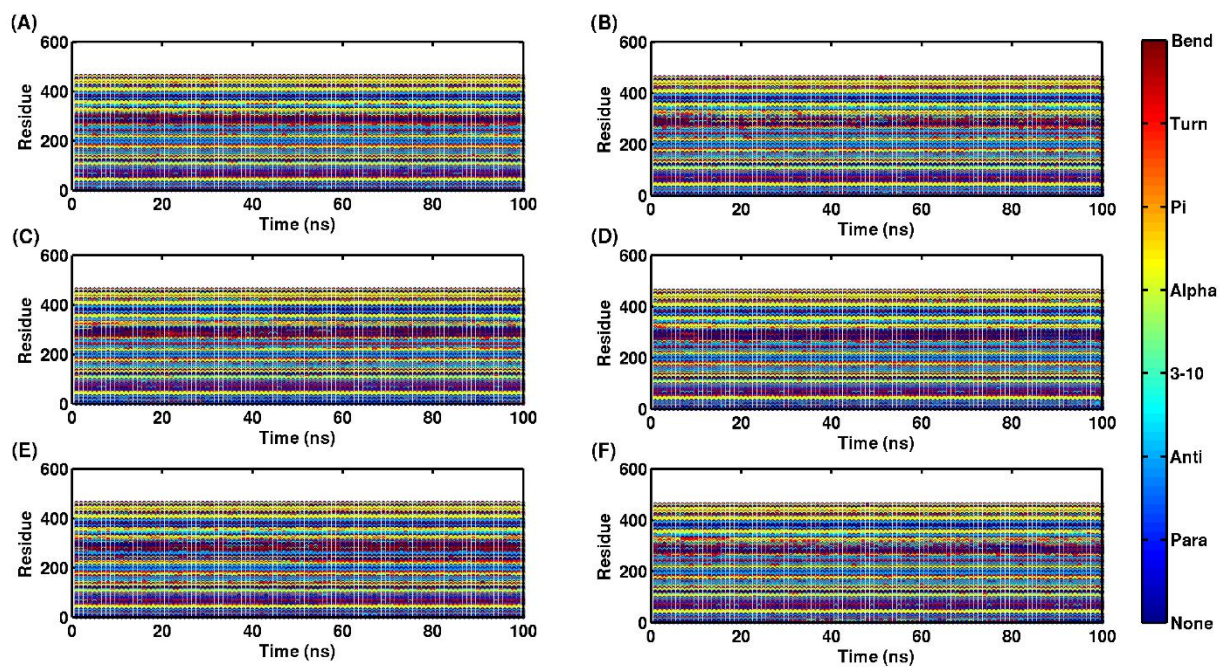


BTZ-N3

Supplementary Figure S3. 2D molecular interaction representation of *DprE1* inhibitors at the active site of *DprE1*.



Supplementary Figure S4. Molecular interaction of compound *C5* at the active site of *DprE1*.



Supplementary Figure S7. The time change in secondary structure of (A) DprE1 (B) DprE1-CT319 (C) DprE1-C5 (D) DprE1-C6 (E) DprE1-C8 (F) DprE1-C10 during the simulation period of 0-100 ns at temperature 300 K, using DSSP.