Research Article

Pharmacophore modelling, QSAR study, molecular docking and insilico ADME prediction of 1,2,3-triazole and pyrazolopyridones as DprE1 inhibitor antitubercular agents

Debadash Panigrahi[1] · Amiyakanta Mishra[1] · Susanta Kumar Sahu[2]

Abstract

Tuberculosis (TB) is a chronic lung infected airborne disease caused by *Mycobacterium tuberculosis* (MTB). The development of resistance towards available antitubercular agents leads to the discovery of new drugs for treatment against these resistant bacteria. Decaprenyl phosphoryl-β-D-Ribose 20-epimerase (DprE1) is a vulnerable target for the design of antitubercular agents which are more acting against multidrug resistant bacterial pathogens. DprE1 is an oxidase involved in the synthesis of arabinogalactan. Inhibition of DprE1 leads to blocking off cell wall synthesis, causing the death of the bacteria. A series of 50 DprE1 inhibitors having activity were subjected to 2D, 3D QSAR, Pharmacophore Modeling, Molecular Docking and in silico ADME studies. Prediction of preliminary Pharmacokinetic and the Drug Likeliness profile was performed for these compounds by *in silico* ADME study. 2D-QSAR and 3D- QSAR models developed by Partial Least Square associated with the Sphere Exclusion method (PLS-SE) and StepWise variable selection method (SW-kNN MFA) based on k- Nearest Neighbor technique are more significant which have cross-validated squared correlation coefficient (q2), coefficient of determination (r2), Fisher ratio (F) values as 0.7499, 0.8917 and 85.04 and the internal (q2 = 0.8198), external (pred_r2 = 0.6109) model validation correctly predicts activity ~ 81% and ~ 61% for the training and test set, respectively. Pharmacophore model was developed with two aromatic regions (Aro), one aliphatic (Ala) and one hydrogen donor (HDr). Docking studies of the selected inhibitors with the active site of DprE1 enzyme showed hydrogen bond interaction with Gly-116, His-131, Arg-118, Thr-117 and Gln-299 residues present at the active site. The results of the present work provide more useful information and important structural insights for designing DprE1 inhibitors with much more enhanced potency.

Keywords DprE1inhibitors · Pharmacophore modelling · 2D · 3D QSAR · Molecular docking · Homology modelling · *Mycobacterium tuberculosis* · ADMET

Abbreviations

DprE1	Decaprenyl phosphoryl-β-D-ribose 20-epimerase enzyme	PLS	Partial least square
		RMSD	Root mean square deviation
MIC	Minimum inhibitoryconcentration	RMSE	Root mean square error
2D-QSAR	Two dimensional quantitative structure-activity relationship	SD	Standard deviation
		r2	Coefficient of determination
3D-QSAR	Three dimensional quantitative structure–activity relationship	q2	Cross validated squared correlation coefficient
		pred_r2	External predictivity

✉ Debadash Panigrahi, debmpharm@yahoo.co.in | [1]Drug Research Laboratory, Nodal Research Centre, College of Pharmaceutical Sciences, Baliguali, Puri- Konark Marine Drive road, Puri, Odisha 752002, India. [2]Department of Pharmacy, Utkal University, VaniVihar, Bhubaneswar, Odisha 751004, India.

k-NNM	K- Nearest neighbour method
F-test	The F test value is the degree of statistical confidence. In general, a QSAR model is more significant as a predictive tool when the higher value is obtained for its F-test
Z score	It can be defined as the absolute difference between the value of the model and the activity field, divided by the square root of the mean square error of the data set. Any compound which shows a value of Z-score higher than 2.5, during the generation of a particular QSAR model is considered as outlier
Aro	Aromatic region
Ala	Aliphatic region
HDr	Hydrogen donor

1 Introduction

Tuberculosis (TB) is an airborne contagious disease caused by *Mycobacterium tuberculosis* and affects about one-third of the world's population [1]. According to the World Health Organization (WHO) tuberculosis continues to cause considerable morbidity and mortality worldwide despite the availability of an effective and economical drug regimen [2–5]. With the emergence and spread of Multi-Drug Resistant Tuberculosis (MDR-TB) [6, 7], Extensively Drug Resistant Tuberculosis (XDR-TB) and deadly complication of tuberculosis infection with Human Immunodeficiency Virus (HIV) demands discovery and development of new antitubercular agents with good efficacy, effectivity and safety, focused on new drug targets with innovative mechanisms of action [8–10].

There are usually three reasons for needing new antituberculosis drugs: (i) To improve current treatment by shortening the total duration of treatment and/or by providing more widely spaced intermittent treatment [11, 12], (ii) To improve the treatment of MDR-TB, and (3) To provide more effective treatment of Latent tuberculosis infection (LTBI) [13].

Computational approaches are made to design DprE1 inhibitors which will target the recently identified enzyme Decaprenyl-phosphoribose 2'-oxidase (DprE1), catalyzes an essential step in mycobacterial cell wall metabolism [14, 15]. The cell wall is a functional and protective interface between the external and internal environments for every living organism. Disruption or inhibition in its synthesis prevents the growth and multiplication of the organism. *Mycobacterium tuberculosis* (MTB) have a special cell wall arrangement, with layers of outer lipids, mycolic acid, polysaccharides (arabinogalactan), peptidoglycan, plasma membrane, lipoarabinomannan (LAM), and phosphatidyl inositol mannoside. The polysaccharides arabinogalactan are the basic precursor for bacterial cell wall synthesis. Deca-prenyl Phosphoryl-β-D-ribose 20-Epimerase (DprE1) is an oxidase enzyme involved in the biosynthesis of Decaprenyl Phosphoryl-D-Arabinose (DPA) which acts as a donor of D-arabino furanosyl residues for the synthesis of Arabinogalactan [16, 17]. DprE1 is a flavoprotein that along with Deca-prenylphosphoryl-2-keto-ribose reductase (DprE2) catalyses epimerization of Decaprenylphosphoryl- D-ribose (DPR) to convert Decaprenylphosphoryl-D-arabinose (DPA) through an intermediate Decaprenylphosphoryl-2-Keto-ribose (DPX) (Fig. 1). This NADP dependent enzymatic reaction makes DprE1 an essential component for cell growth and survival of bacteria [18–20]. Hence DprE1 is evolved as an important drug target, inhibition of this enzyme block bacterial cell wall synthesis leads to the death of the bacteria [21].

Computational approaches are made to develop the new active and minimum toxic drug moieties. Computational modelling of drugs is based on information about the ligand and the target receptor [22]. Based on published information about ligands and receptor, either Structure based or Ligand-based Molecular Design approaches are used to correlate the biological activity with the chemical structure of the ligands [23]. Computational studies are considered effective tools in medicinal chemistry and are useful in speeding up the drug design process [24].

Molecular modelling represents the molecular structure numerically and their activity as the equation of quantum. An attempt is made in the present work to perform Quantitative Structure Activity Relationship (QSAR) study, Pharmacophore Modelling, Molecular Docking and insilico ADME prediction on a series of DprE1 inhibitors. QSAR study includes development of two-dimensional (2D) and three dimensional (3D) models where the structure of the molecules taken in the most stable state which are using to calculate the descriptors [25]. Validations of the developed models are carried out using different statistical parameters [26]. The validated models developed in this study help to optimize the lead compounds and provide information about the correlation between structural properties and activity [27]. Pharmacophore Modelling and Molecular Docking study are performed to understand and to interpret the binding interactions mechanism between the ligands and the receptor. Insilico ADMET prediction of drug molecules help to assess the Pharmacokinetic (PK) profile and drug likeliness of molecules [28]. The composition of Docking and Pharmacophore Modelling with QSAR studies can be applied to gain more precise information on the interactions between the ligand and the receptor [29–31].

Based on the developed models, rational design of novel active DprE1 inhibitors are made which are having greater selective, effective and safety, therapeutic activity. The got results of Pharmacophore and Docking study can improve the binding process of ligands with its receptor and provide insights into the structural features related to the activities of the new drug compounds.

Fig. 1 NAD dependent Biochemical reaction catalysed by DprE1 and DprE2 enzymes in Mycobacterium
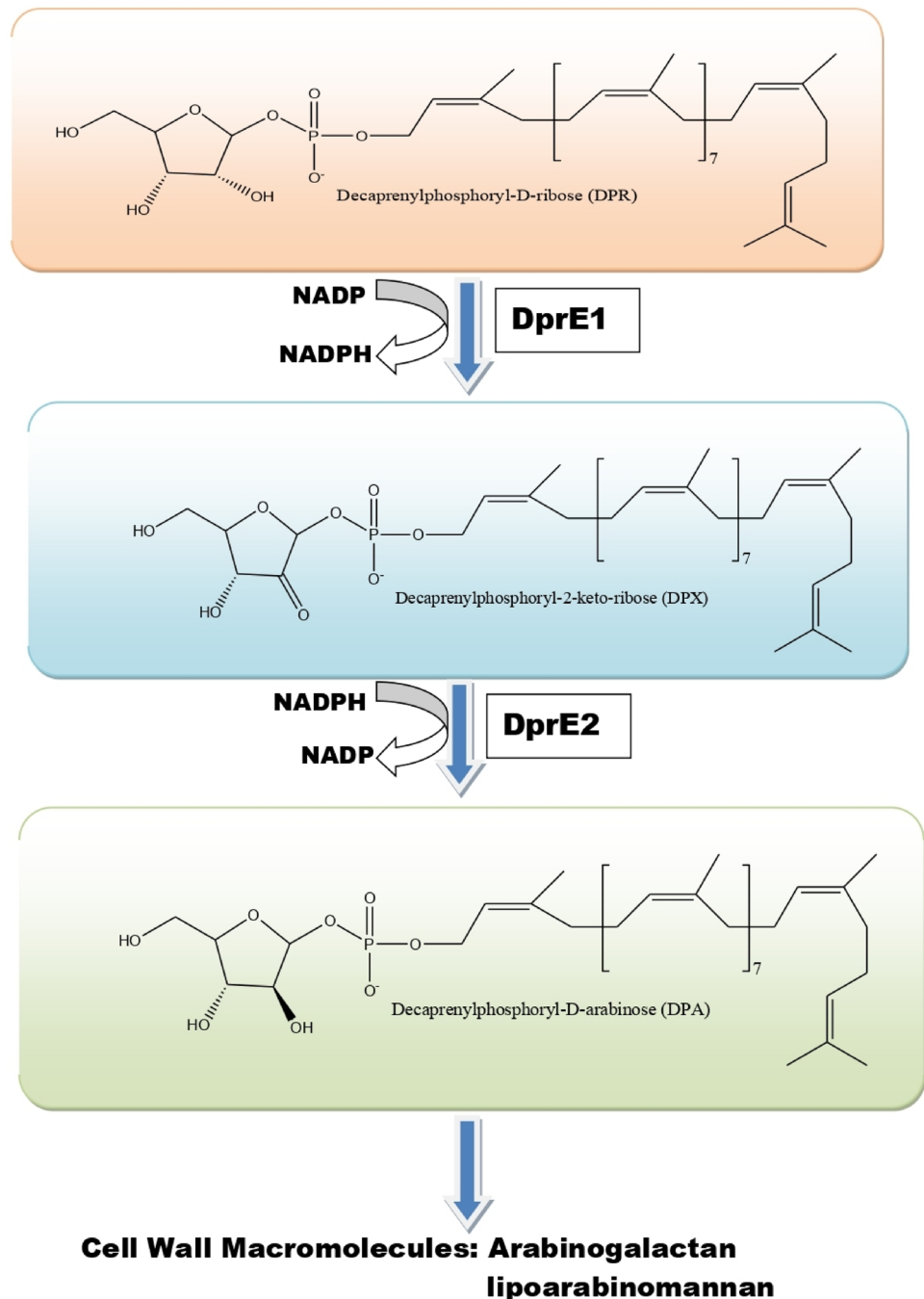
2 Materials and methods

2.1 Data set

To perform the present computational study, a set of 50 compounds having reported IC^{50} values were taken from the available literature [32, 33] excluding compounds having not well defined biological activities. The selected compounds for the study shared the same activity and assay procedure with significant variations in their structure

and potency. Inhibitory potencies of the compounds in the data set have IC^{50} values ranges from 0.005 to 56.7 µm which were further converted to pIC^{50} by using the following mathematical formula given as Eq. 1;

$$pIC^{50} = -\log_{10}\left(IC^{50}\right) \qquad (1)$$

The structure of all the compounds given in the data set is sketched using the molecular sketching facilities provided in the MDS software of V-Life [34]. Energy

minimization of the compounds is done by using Merck molecular force field (MMFF) [35, 36] using MDS software of V-Life by fixing a dielectric constant at 1.0 and root mean square (RMS) gradient at 0.0001. Energy minimization of the compounds is made for effective binding of the drug with its target receptor. The division of whole data set into training and test sets is based on Sphere Exclusion Algorithms, so that the activity of the selected test set are distributed throughout the activity column of the compounds, the distribution curve for test and training compounds is given in Fig. 2. The QSAR models are developed and validated by taking 36 and 14 molecules as training and test set compounds. The chemical structure and their pIC^{50} values are given in Table 1.

2.2 QSAR study

QSAR study is performed to find the correlation between the activity and structural features (descriptors) of the data set molecules. In this method, we try to find structural parameters that relate to the inhibition activity through mathematical equations [37, 38].

2.2.1 Generation of 2D-QSAR models

In, the present study 2D QSAR models are developed between activity and descriptors like Retention Index (chi),Atomic valence connectivity index (chiv), Path Count, Chi Chain, Path Count, Path Cluster, Element Count, Dipole Moment, topological, Estate Contributions, Information Theory Index, Extended Topochemical Atom (ETA) based descriptors, Polar Surface Area etc. consider as Physiological descriptors, T_2_O_7, T_2_N_5, T_N_N_5, T_2_2_6, T_C_O_1, T_O_Cl_5 etc. as Alignment Independent (AI) and MMFF atom types descriptors. 3D descriptors such as Electro Static, Distance Based Topological Indices, SemiEmpirical and Hydrophobicity base logP descriptors are excluded during the study. 415 molecular descriptors are calculated using V-Life MDS software before developing QSAR models. For alignment independent descriptors, we have used attributes (2, T, C, N, O, F, S, Cl) range from 0 to 7 and structure descriptors as Topological in the software. After obtaining the values of descriptors for all the compounds, descriptors that have a constant value for all the molecules are discarded. Four QSAR models are developed by using Multiple Regression (MR), Principal Component Regression (PCR), Partial Least Square Regression (PLSR) and Partial Least Square associated with the Sphere Exclusion (PLS-SE) methods taking all the calculated descriptors as independent variables and biological activity as the dependent variable.

2.3 Generation of 3D-QSAR models

3D QSAR models for the above data set are developed by using k-Nearest Neighbour Molecular Field Analysis (kNN- MFA) principle. The values of the 3D descriptors such as Electrostatic and Steric parameters are calculated by setting the dielectric constant as 1.0, charge type as Gasteiger-Marsili and a sp3 carbon probe atom with charge 1.0. The cut off energy of 10.0 kcal/mol and 30 kcal/mol are set as the default for electrostatic and steric energies. A total of 2080 field descriptors (1040 for each electrostatic and steric) are calculated using MDS software. 3D QSAR models are developed by setting a cross-correlation limit as 0.5, the number of variables in the equation as 4, term selection criteria as q2, F-test in and out value as 4 and 3.99 respectively. Three models are developed by Step Wise variable Selection Method (SW-kNN MFA), Simulated Annealing Variable Selection Method (SA-kNN MFA) and Genetic Algorithm Variable Selection Method (GA-kNN MFA).

2.3.1 Model validation

For validation of the developed QSAR models, the data set is divided into two sets as training and test sets. This division is based on the substitution groups and the inhibition of compounds. The training set is employed to produce the QSAR model, and the test set is used to validate the quality of the developed models. The statistical parameters of the developed models, internal and external validations are adopted for testing the fitness, stability and predictive ability of the QSAR models. Both the developed 2D and 3D QSAR models are validated by considering many statistical parameters such as the number of compounds in regression (n), the number of variables (k), degree of freedom, squared correlation coefficient (r2), cross-validated correlation coefficient (q2), Fischer's value (F test) and r2 for external test set, (pred_r2) for external validation. For
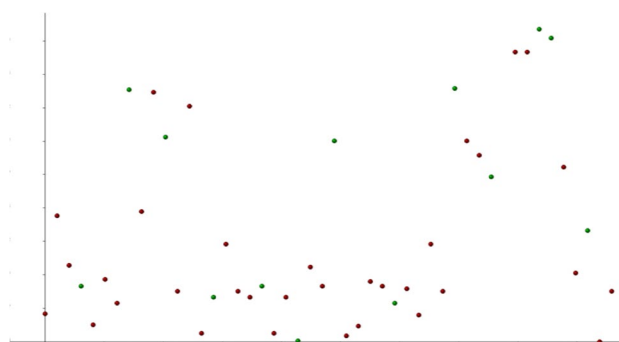

Fig. 2 Distribution curve of Test (Green Dot) and Training set (Red Dot) compounds

Table 1 Chemical structure and pIC^{50} values of the compounds having DprE1 inhibition activity
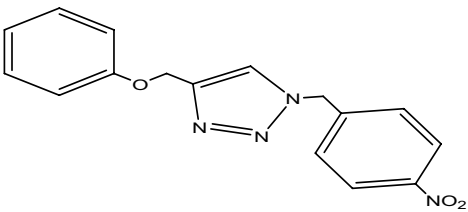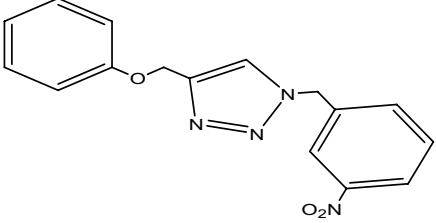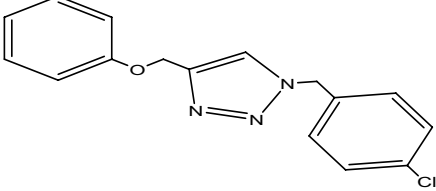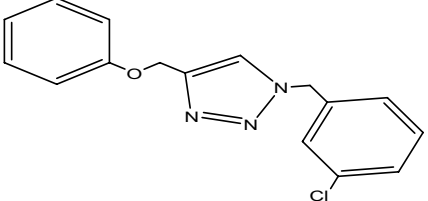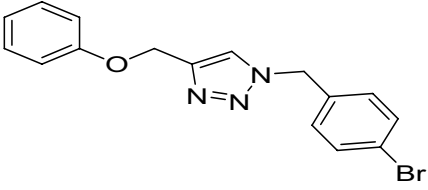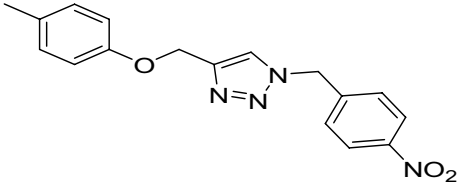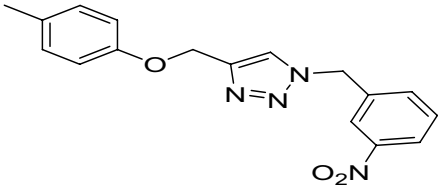
S. no	Structure of compound	QSAR set	pIC^{50}
1		Test Set	0.231
2		Test Set	0.381
3		Test Set	0.545
4		Training Set	0.381
5		Test Set	0.506
6		Training Set	0.042
7		Training Set	0.381

Table 1 (continued)
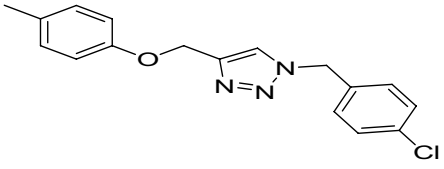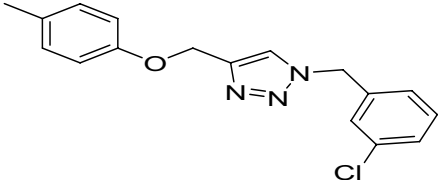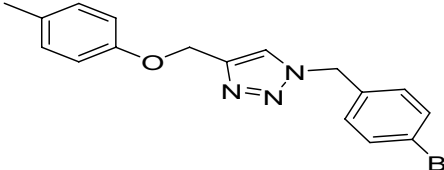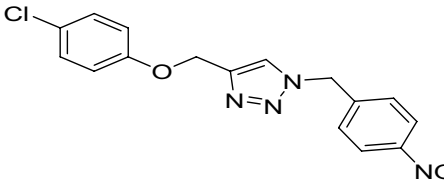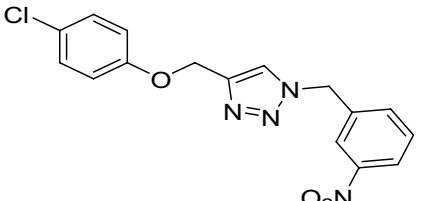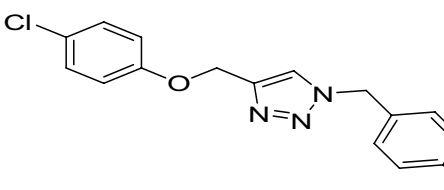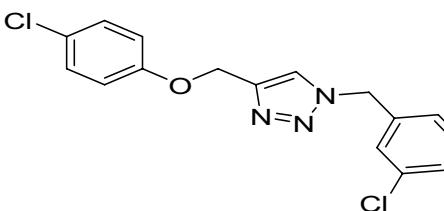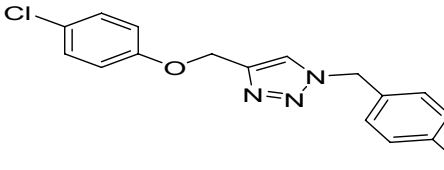
S. no	Structure of compound	QSAR set	pIC^{50}
8		Training Set	0.398
9		Test Set	0.478
10		Training Set	0.893
11		Training Set	0.554
12		Training Set	0.415
13		Training Set	0.155
14		Test Set	0.463
15		Test Set	0.302

Table 1 (continued)
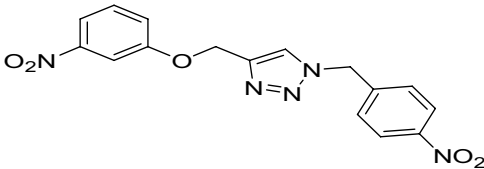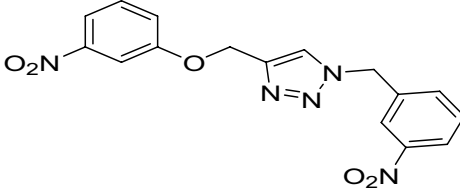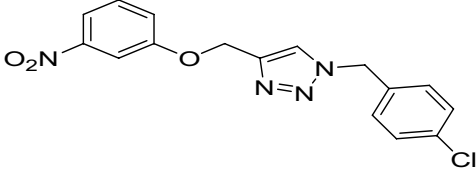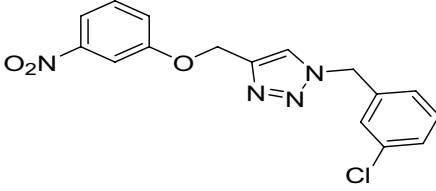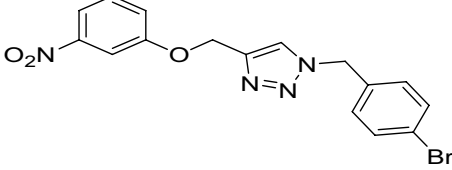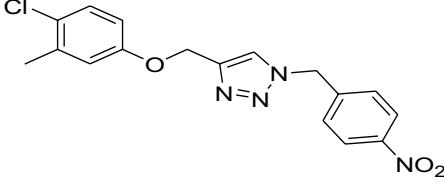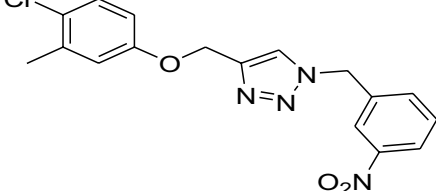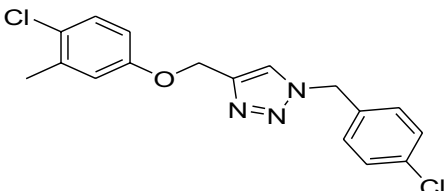
S. no	Structure of compound	QSAR set	pIC^{50}
16		Training Set	1.748
17		Training Set	0.920
18		Training Set	1.728
19		Test Set	1.425
20		Training Set	1.636
21		Test Set	0.097
22		Test Set	0.343
23		Training Set	0.699

Table 1 (continued)
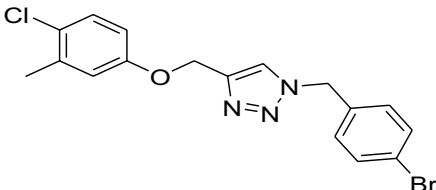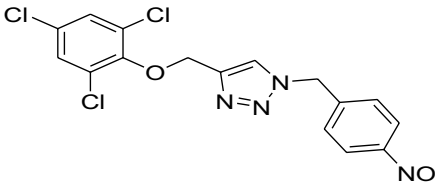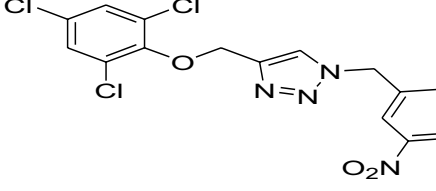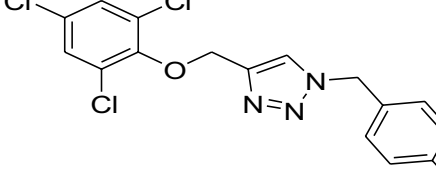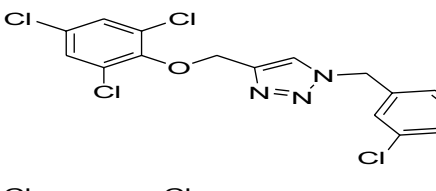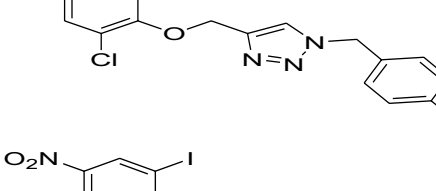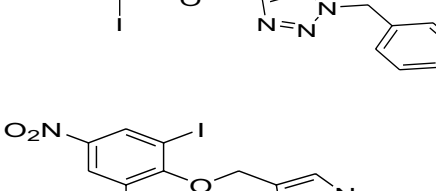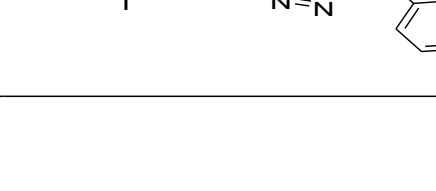
S. no	Structure of compound	QSAR set	pIC^{50}
24		Training Set	0.381
25		Training Set	0.343
26		Training Set	0.415
27		Training Set	0.097
28		Test Set	0.343
29		Training Set	0.046
30		Training Set	0.415
31		Training Set	1.398

Table 1 (continued)

S. no	Structure of compound	QSAR set	pIC^{50}
32		Test Set	0.080
33		Test Set	0.147
34		Training Set	0.448
35		Training Set	0.415
36		Training Set	0.302
37		Training Set	0.398
38		Test Set	0.222
39		Training Set	0.699
40		Training Set	1.754
41		Training Set	1.398

Table 1 (continued)

S. no	Structure of compound	QSAR set	pIC^{50}
42		Training Set	1.302
43		Training Set	1.155
44		Training Set	2.302
45		Training Set	2
46		Training Set	2
47		Training Set	2.155
48		Training Set	2.097
49		Training Set	1.222
50		Training Set	0.793

the internal predictive ability of the model Leave One Out (LOO) method is used showed as the value of q2 (cross-validated explained variance) [39].

External validation of the developed QSAR models is performed by measuring the predictive power of the current models on the external test set by calculating the pred_r2 value as given in Eq. 2, which gives the statistical correlation between predicted and actual activities of the test set compounds.

$$\text{pred}_{r2} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \qquad (2)$$

where $y_i, \hat{y}_i$ and y_{mean} are the actual, predicted activity of the ith molecule in the test set and the average activity of all the molecules in the test set, respectively.

Internal validation of the developed QSAR models is performed by calculating the q^2 value as given in Eq. 3, which gives the statistical correlation between predicted and actual activities of the training set compounds.

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \qquad (3)$$

where $y_i, \hat{y}_i$ and y_{mean} are the actual, predicted activity of the ith molecule in the training set and the average activity of all the molecules in the training set, respectively.

2.4 Pharmacophore generation

The development of pharmacophore model is one of the important tasks in drug design and bioactivity prediction. A pharmacophore model has described a set of three-dimensional features which are necessary for bioactive ligands [40, 41]. It shows about the nature of the functional groups like hydrogen bond donors, acceptors, hydrophobic areas, charge interactions, non-covalent bonding and interchange distances which affect the ligand- target interactions. The MolSign module in VLifeMDS provides tools for aligning small organic molecules based on their three dimensional pharmacophore features. Pharmacophore modelling is performed by taking TCA-1 as the reference compound and all 50 compounds for alignment. The primary pharmacophore feature count, enter the value 4, shows the minimum number of pharmacophore features generated for an alignment. The tolerance field, enter the value 10 Å, shows the flexibility in percentage allowed while comparing two feature combinations across two molecules. The Max Distance allowed between the two features, set as 10.

2.5 Molecular docking

Molecular docking study is a computational approach for searching a ligand that can fit both geometrically and energetically into the binding site of a target to show action. Docking study helps to predict drug/ ligand or receptor/ protein interactions by identifying the suitable active sites in the protein, getting the best geometry of ligand-receptor complex and calculating the energy of interaction for different ligands to design more effective ligands [42, 43]. In the present work, docking study is performed for all 50 compounds with the DprE1 enzyme. The whole study is carried out by the Biopredicta tool of V-Life MDS software version 4.6. X-ray diffraction crystal structure of *M. Tuberculosis* DprE1 is obtained from RCSB Protein data bank (PDB ID-4KW5) complex with inhibitor TCA-1, having resolution 2.612 A^0 is used for docking study. Initially, the enzyme is bound with a ligand (TCA-1, dock score − 3.453), which is removed and the missing loops are added with the help of homology modeling modules of the software. During study bond orders of the ligands are assigned, hydrogen atoms are added and the water molecules which do not involve in the interaction are deleted. The TCA-1 bound cavity is considered carrying out the docking study of the selected 50 compounds. Finally, the best-docked structures are selected using their dock score. The interacting amino acids are identified as Val-120, Thr-117, Arg-57, Pre-117, Gly-116, His-131, Ser-122, Tyr-284 and Lys-330 present in the binding site of the target enzyme.

2.6 Drug likeliness and in silico ADME prediction

Earlier prediction of the ADMET properties of drug molecules helps a lot towards drug discovery. This information helps to assess the pharmacokinetic (PK) profile of molecules. The PK properties of the molecules depend on their chemical descriptors, which determine their ADMET. The PK parameters are calculated by ADMET lab a user-friendly freely available web interface [44–46]. Several mathematical predictive models for different PK parameters are available such as Aqueous Solubility, Apparent Caco-2, log Kp for skin permeability, Blood–brain barrier (BBB), Volume of Distribution (Vd), Plasma Protein Binding, Metabolism, Elimination {Half lifetime ($T_{1/2}$), Clearance rate (CL)} and Toxicity which are used to predict the ADMET properties of the drug molecules. The drug-likeness (DL) analysis module includes five commonly used drug- likeness rules (Lipinski, Ghose, Oprea, Veber and Varma) and parameters, such as molecular weight (MW) of ≤ 500 amu, a logP value of ≤ 5, hydrogen bond donor ≤ 5 and hydrogen

bond acceptor site (N and O atoms) ≤ 10, the number of rotatable bonds ≤ 10 and topological polar surface area (TPSA) ≤ 140 Å. The significant predicted pharmacokinetic and physicochemical descriptors accounts for druggablilty of a molecule [47, 48].

3 Result and discussion

3.1 Development and validation of 2D-QSAR models

2D QSAR models were developed by considering all the two dimensional calculated descriptors as independent variables and biological activity as the dependent variable. For internal and external validation of the developed models, the data set of the compounds was divided into 14 and 36 as the test and training sets, respectively. The correlation between actual and predicted activity for both training and test set compounds is shown in Table 2. Uni-column statistic is performed for both training and test series to check the spread of data. The results of the uni-column statistics study are presented in Table 3. From the result, it was clear that the test set is interpretive, i.e. the activity of the test set derived within the activity range of the training set. The mean and standard deviation of the training and test sets provides insight into the relative difference of mean and point density distribution of the two sets. As the average of the test set is higher than the training set shows the presence of relatively more active molecules as compared to the inactive ones.

2D QSAR models are developed by using 4 methods multiple regression (MR), principal component regression (PCR), partial least square regression (PLSR) and partial least square associated with the sphere exclusion (PLS-SE), the correlation equations between activity (pIC50) and the selected parameters are given as Eqs. 4, 5, 6 and 7 respectively. Followed by the validation of the developed QSAR models to check both internal and external predictive power, which implies a quantitative assessment of model robustness. Validation of the four developed QSAR models is confirmed based on values for various studied statistical parameters; the result of the study is given in Table 4.

$$pIC_{50} = 0.5850(\pm0.0742)T\_O\_O\_5$$
$$- 0.0524(\pm0.0143) T\_2\_2\_4 + 0.9615 \quad (4)$$

$$pIC_{50} = 0.0000(\pm0.0000) lpc + 0.4412 \quad (5)$$

$$pIC_{50} = 1.1867T\_N\_O\_4 - 0.0174SdOE - index - 1.9355 \quad (6)$$

$$pIC_{50} = -0.2473SssssCE \text{ index} + 1.0914T\_N\_O\_4$$
$$+ 11.5647Most - vePotential - 0.6078SsssNE \text{ index}$$
$$- 0.0162SAHydrophilicArea - 0.0588T\_T\_Cl\_4$$
$$- 0.1418SaaCHcount + 2.094$$
$$(7)$$

From the data given in above table it is clear that the QSAR model developed by Partial Least Square associated with the Sphere Exclusion method (PLS-SE) is statistically more significant than others because the calculating r2 and r2_se for training and the same coefficient for external test set (pred_r2) are having values 0.8917, 0.2407 and 0.5935 with the low standard error of estimation shows overall internal statistical significance level better than 99.9% as the F-test having value 85.0374. This model accounts for 89% variance in the inhibitory activity. The value of the cross-validated Square Correlation Coefficient (q2) is 0.7499 suggesting the good predictive ability of the model. This model shows the interrelationship between the activity and the parameters such as SssssCE-index, T_N_O_4, Most-vePotential, SsssNE-index, SAHydrophilicArea, T_T_Cl_4 and SaaCHcount, contribution plot of these parameters towards activity is presented in Fig. 3. The positive coefficient of T_N_O_4 and Most-vePotential shows that antitubercular activity will increase with the increase in the number of Nitrogen atoms separated from Oxygen atom by 4 bonds and increase the -ve potential in the Vander Waals surface area of the molecule. Whereas the negative coefficient for the parameters SssssCE-index, SsssNE-index, SAHydrophilic Area, T_T_Cl_4 and SaaCH count shows the activity will increase with the decrease in eletrotopological state indices for the number of carbon atom and –NH group connected with 4 and 3 single bonds respectively, vdW surface descriptors showing hydrophilic surface area, the number of chlorine atom separated by 4 bonds and total number of carbon atoms connected with a hydrogen along with 2 aromatic bonds.

The fitness plot between actual and predicted activity for training and test set compounds given in Fig. 4 provides an idea about how well this model is trained and how well it predicts the activity of the external test set. Further, the distribution curve of actual and predicted activity for training and test sets compounds for the well-developed model are represented in Fig. 5a, b, depicting closeness between the actual and predicted activity of the compounds for training and test set.

3.2 Development and validation of 3D-QSAR models

By using k-Nearest Neighbour Molecular Field Analysis (kNN- MFA) principle 3D- QSAR models for the above data set are developed. Three models are developed by Step

Table 2 Observed and predicted activities (pIC50) for the Training and Test set compounds

Compound no.	2D- QSAR model (PLS-SE)			3D- QSAR model (SW-kNN MFA)		
	Experimental pIC^{50}	Predicted pIC^{50}	Residual activity	Experimental pIC^{50}	Predicted pIC^{50}	Residual activity
1	0.231	0.265	− 0.034	0.231	0.256	− 0.025
2	0.381	0.514	− 0.133	0.381	0.432	− 0.051
3	0.545	0.541	0.004	0.545	0.586	− 0.041
4	0.381	0.415	− 0.034	0.381	0.421	− 0.040
5	0.506	0.601	− 0.095	0.506	0.595	− 0.089
6	0.042	0.087	− 0.045	0.042	0.067	− 0.025
7	0.381	0.451	− 0.070	0.381	0.376	0.005
8	0.398	0.385	0.013	0.398	0.336	0.062
9	0.478	0.485	− 0.007	0.478	0.468	0.010
10	0.893	0.763	0.130	0.893	0.723	0.170
11	0.554	0.493	0.061	0.554	0.598	− 0.044
12	0.415	0.423	− 0.008	0.415	0.534	− 0.119
13	0.155	0.174	− 0.019	0.155	0.203	− 0.048
14	0.463	0.486	− 0.023	0.463	0.627	− 0.164
15	0.302	0.395	− 0.093	0.302	0.376	− 0.074
16	1.748	1.652	0.096	1.748	1.564	0.184
17	0.920	0.894	0.026	0.920	1.045	− 0.125
18	1.728	1.685	0.043	1.728	1.612	0.116
19	1.425	1.365	0.060	1.425	1.336	0.089
20	1.636	1.607	0.029	1.636	1.538	0.098
21	0.097	0.159	− 0.062	0.097	0.148	− 0.051
22	0.343	0.388	− 0.045	0.343	0.395	− 0.052
23	0.699	0.654	0.045	0.699	0.610	0.089
24	0.381	0.471	− 0.090	0.381	0.487	− 0.106
25	0.343	0.452	− 0.109	0.343	0.443	− 0.100
26	0.415	0.512	− 0.097	0.415	0.567	− 0.1520
27	0.097	0.174	− 0.077	0.097	0.132	− 0.035
28	0.343	0.396	− 0.053	0.343	0.468	− 0.125
29	0.046	0.125	− 0.079	0.046	0.145	− 0.099
30	0.415	0.507	− 0.092	0.415	0.556	− 0.141
31	1.398	1.263	0.135	1.398	1.211	0.187
32	0.080	0.174	− 0.094	0.080	0.155	− 0.075
33	0.147	0.197	− 0.050	0.147	0.213	− 0.066
34	0.448	0.528	− 0.080	0.448	0.598	− 0.150
35	0.415	0.593	− 0.178	0.415	0.574	− 0.159
36	0.302	0.417	− 0.115	0.302	0.454	− 0.152
37	0.398	0.317	0.081	0.398	0.287	0.111
38	0.222	0.238	− 0.016	0.222	0.365	− 0.143
39	0.699	0.596	0.103	0.699	0.625	0.074
40	1.754	1.535	0.219	1.754	1.606	0.148
41	1.398	1.285	0.113	1.398	1.176	0.222
42	1.302	1.258	0.044	1.302	1.236	0.066
43	1.155	1.214	− 0.059	1.155	1.245	− 0.090
44	2.302	1.958	0.344	2.302	2.176	0.176
45	2	1.977	0.023	2	1.887	0.113
46	2	1.839	0.161	2	1.764	0.236
47	2.155	1.857	0.298	2.155	2.378	− 0.223
48	2.097	2.011	0.086	2.097	1.987	0.110

Table 2 (continued)

Compound no.	2D- QSAR model (PLS-SE)			3D- QSAR model (SW-kNN MFA)		
	Experimental pIC^{50}	Predicted pIC^{50}	Residual activity	Experimental pIC^{50}	Predicted pIC^{50}	Residual activity
49	1.222	1.198	0.024	1.222	1.269	−0.047
50	0.793	0.758	0.035	0.793	0.657	0.136

Table 3 Unicolumn statistics of activity (pIC50) for Training and Test set compounds for 2D-QSAR

Compounds	Average	Maximum	Minimum	Std. Dev	Sum
Training set	0.8897	2.3010	0.0414	0.7137	32.0306
Test set	0.9699	1.4249	0.0969	0.3720	6.9979

Wise variable Selection Method (SW-kNN MFA), Simulated Annealing variable Selection Method (SA-kNN MFA) and Genetic Algorithm variable Selection Method (GA-kNN MFA) by considering 3D descriptors such as Electrostatic and Steric parameters. The QSAR models for all three methods are given in Eqs. 8, 9 and 10, respectively. To check the predictivity of the developed models, the data set is divided into training and the test set with 34 and 16 compounds. The correlation between actual and predicted activity for both training and test set compounds is shown in Table 2.

$$pIC_{50} = E\_698(-6.1424 - 5.7807)$$
$$+ E\_225(6.7907 - 7.1363)$$
$$+ S\_532(-0.4929 - 0.4784) \quad (8)$$

$$pIC_{50} = S\_412(-0.6886 - 0.3326) + E\_276(-2.2760 - 0.7668)$$
$$+ E\_165(1.8097 - 2.9967) + E\_585(-8.0366 - 0.8839) \quad (9)$$

$$pIC_{50} = E\_207(-9.0489 - 6.7334) + S\_448(30.0000)$$
$$+ E\_682(-1.0329 - 0.8132)$$
$$(10)$$

Unicolumn statistic study is performed on training and test sets, the result is in Table 5, which signifies that test set contains more active molecules and is uniformly distributed within the min–max range of the training set.

Validation of the three developed models is performed to determine the best model that correlates the activity with the descriptors. The result of the validation study is given in Table 6.

The validation study result of the developed 3D- QSAR models suggests that the model developed by SW-kNN MFA method given in Eq. 8 is statistically more significant and better than other two regarding the internal ($q2 = 0.8198$) and the external ($pred\_r2 = 0.6109$) predictive, shows predict ability of ~82% and ~61% for the training and test set, respectively. This model shows that the contributing descriptors are E_698, E_225 and S_532 spread along as field points, the correlation plot is shown in Fig. 6. Electrostatic fields at E_698 ($-6.1424 - 5.7807$) and E_225 ($6.7907 - 7.1363$) are in the negative and positive range near to ring towards activity showing substitution of electronegative and electropositive groups in these sites enhances the activity. Further negative coefficient of the steric factor at S_532 ($-0.4929 - 0.4784$) shows the substitution of a less bulky group in this region is preferable for the increase of activity. The fitness plot between the

Table 4 Statistical validation results of the developed 2D-QSAR models

Method	Degree of freedom	r^2	q^2	F_test	r^2_se	q^2_se	pred_r^2	pred_r^2se
MR	32	0.8165	0.7556	47.4689	0.2438	0.2813	0.4853	0.6871
PCR	34	0.4845	0.4439	31.9547	0.3964	0.4117	0.2965	0.8034
PLSR	33	0.8158	0.7184	73.0779	0.2405	0.2974	0.3819	0.7530
PLS-SE	31	0.8917	0.7499	85.0374	0.2407	0.3657	0.5935	0.4483
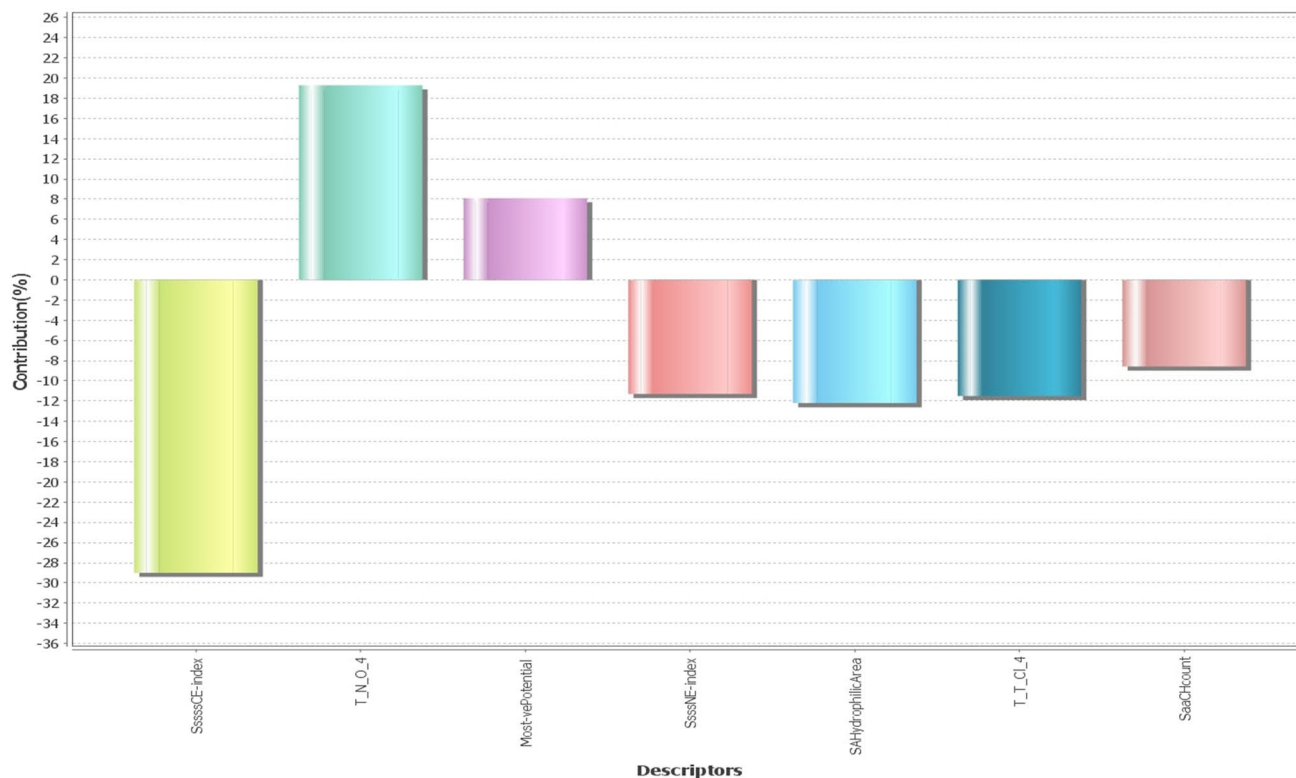
Fig. 3 Contribution plot of parameters towards activity

actual and predicted activity of the developed model for training and test set compounds is shown in Fig. 7 which provides an idea about its good predictivity. The distribution curve of actual and predicted activity for training and test set compounds is given in Fig. 8a and b.

3.3 Pharmacophore modelling

In the present work, Pharmacophore modelling for all the compounds present in the series is carried out by taking TCA-1 as a reference compound. Pharmacophore modelling provides useful information to design and synthesise novel potent DprE1 inhibitors. Pharmacophore model is developed by taking four necessary features for the activity of ligand, the results are shown in Fig. 9a, b and Table 7.The obtained Pharmacophore model contains two

aromatic (Aro) centre (Yellow sphere), one aliphatic (Ala) carbon centre (Orange sphere) and one hydrogen bond donor (Hdr) centre (Green sphere) reveals that these features are necessary for showing DprE1 inhibiting activity.

3.4 Molecular docking studies

Molecular Docking study is carried out for all 50 compounds with the binding site of the target DprE1 enzyme. The grid docking score values of all compounds are given in Table. 8. Based on the grid dock score, five compounds of number 8,15,16,27 and 35 are selected for the study showing good binding efficiency with the target enzyme. The binding modes of these compounds are given in Fig. 10a–e respectively. Docking study reveals that these molecules are interacting with amino acid residues like

Fig. 4 Fitness Plot for 2D QSAR model developed by PLS-SE method
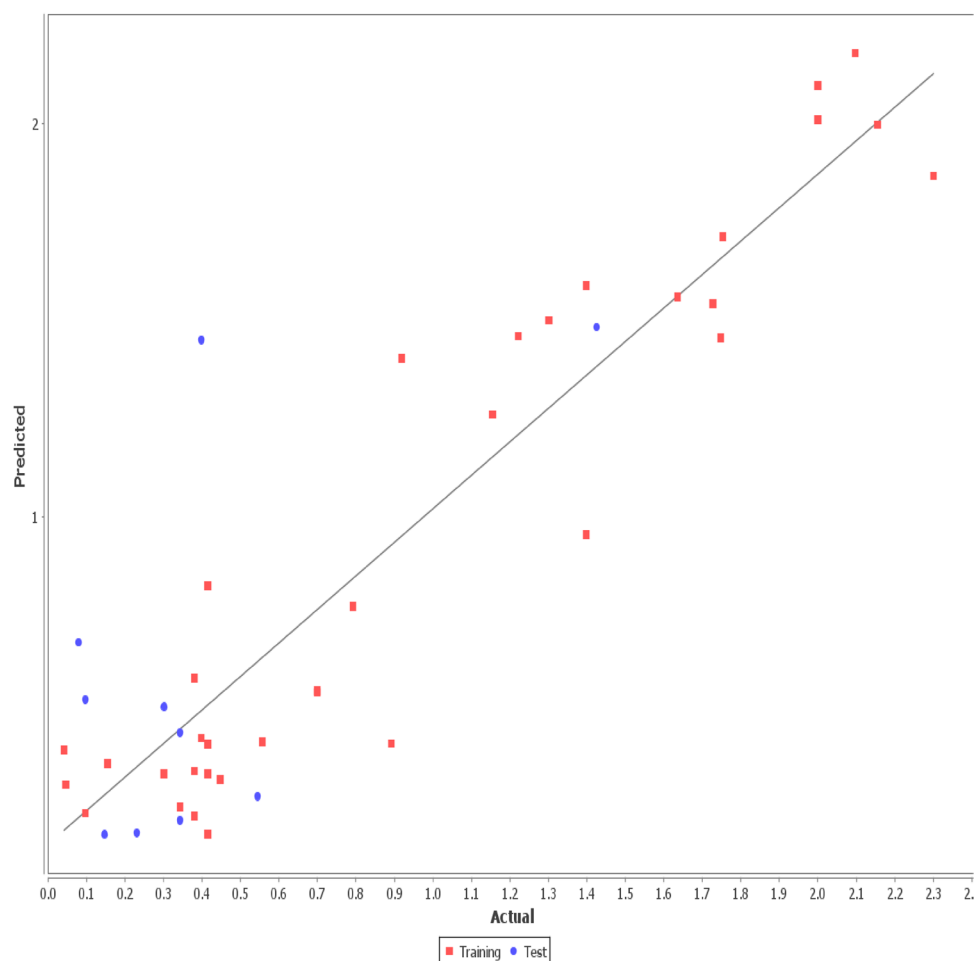
Gly-116, His-131, Arg-118, Thr-117 and Gln-299 present at the active site of the target enzyme by forming H-bond with them. The two dimensional binding representation of these compounds with the target enzyme are given in Fig. 11a–e respectively shows the interaction of these compounds with active site amino acids. Two-dimensional ligand interaction plot of these compounds are shown non polar interaction because of the formation of hydrogen bonds (H-bond) between amino acids and atoms (O and N) present in the chemical structure of these compounds, the interaction result is given in Table 9. Docking study of these molecules with the target site contribute that substitution of electron donating groups on these particular sites increases the binding efficacy by forming H-bond

with the target site and potentiate the DprE1 inhibiting action, hence it help towards the design and development of potent and selective lead molecules having DprE1 inhibiting antitubercular action.

3.5 Drug likeliness, in silico ADME and toxicity study

Knowing ADME features about a compound in advance is important for drug discovery, and poor pharmacokinetics (PK) is the major concern for the failure of drug candidates in clinical trials. Therefore, knowing of ideal ADME properties at earlier stages helps to generate good potential candidates that can avoid the latter stage of elimination

a

b
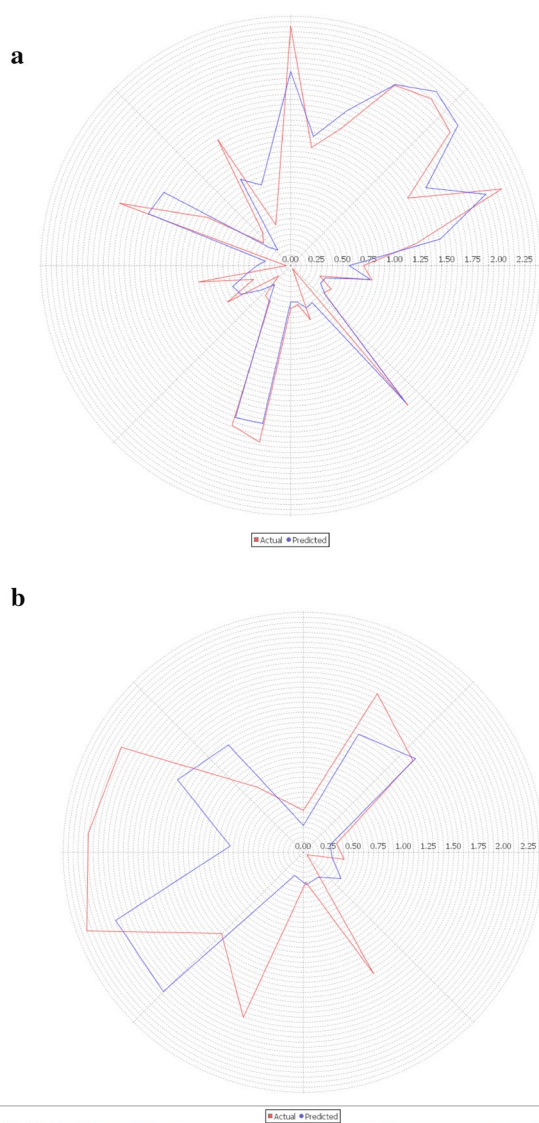
Fig. 5 **a** and **b** Actual Vs Predicted activity of Training and Test set for 2D QSAR developed by PLS-SE method

Table 5 Unicolumn statistics of activity (pIC50) for Training and Test set compounds for 3D QSAR

Compounds	Average	Maximum	Minimum	Std. Dev	Sum
Training set	0.8908	2.3010	0.0458	0.7216	30.2885
Test set	0.5463	1.6355	0.0414	0.4230	8.7400

Fig. 6 Field points exhibiting contributing descriptors for 3D-QSAR model by SW-kNN MFA method

and can easily pass from clinical trial studies. With this aim in the present study, all the 50 compounds present in the series are used for prediction of their Pharmacokinetic (ADME) parameters, drug toxicity, and drug likeliness features by using ADMET lab web interface. The predicted results of Pharmacokinetics, Toxicity and Drug likeliness are presented in Tables 10 and 11, respectively.

The predicted result showed that all the compounds satisfy the Lipinski's rule of five for drug likeliness and oral bioavailability. Values for the distribution coefficient D (LogD) and distribution coefficient P (LogP) are within the optimal range for all the compounds suggest the idealness of these compounds. The Solubility (LogS) values are in optimum range, suggesting good dissolution and absorption of drugs. The optimum values of other descriptors related to absorption suggest good intestinal absorption and skin permeability of these compounds. Optimum values of Topological polar surface area (< 140 Å^2) and rotatable bonds (0–15) holds a great effect towards oral bioavailability of these compounds. The predicted result shows good plasma protein binding, Blood-Brain Barrier

Table 6 Statistical Validation results of the developed 3D-QSAR models

Method	n	Degree of freedom	q^2	q^2_se	pred_r^2	pred_r^{2}se
SW-kNN MFA	34	30	0.8198	0.3063	0.6109	0.3448
SA-kNN MFA	34	29	0.7939	0.3276	0.5295	0.3792
GA-kNN MFA	34	30	0.1755	0.6552	0.4960	0.3925

Fig. 7 Fitness curve for 3D-QSAR model by SW-kNN MFA method
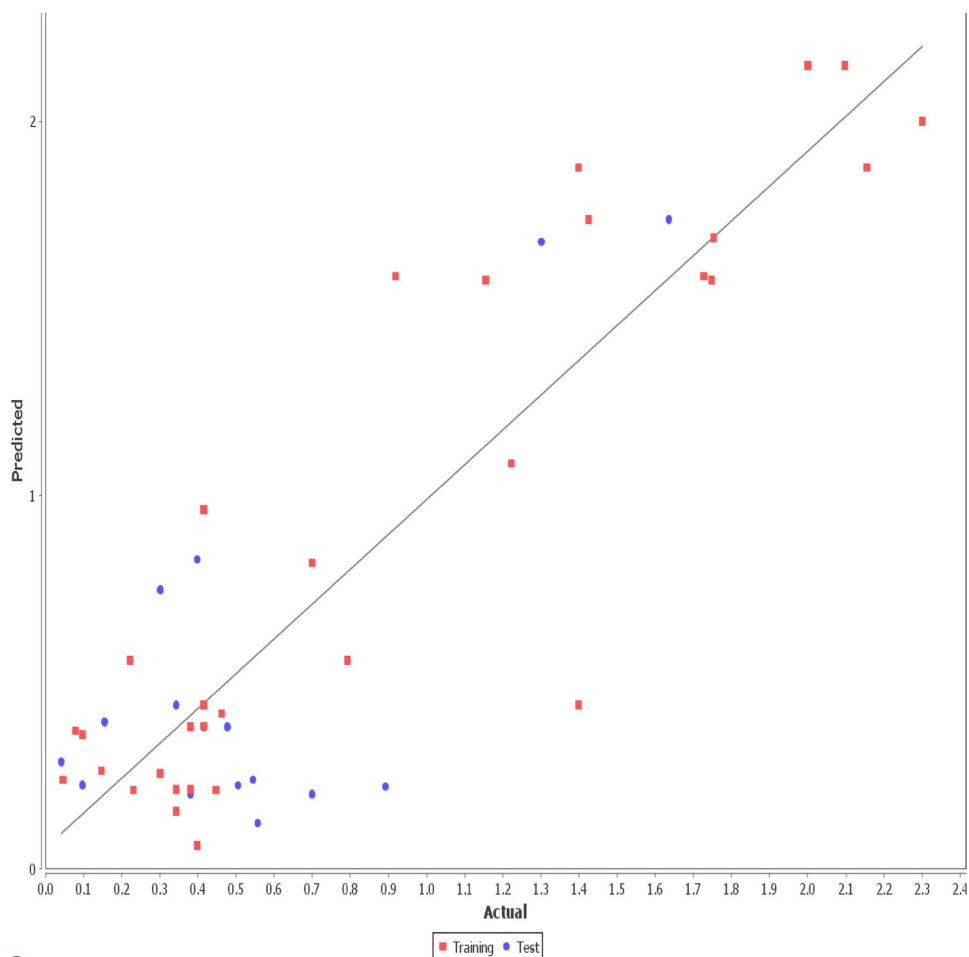
penetration (BBB) ability, low half-life (T1/2) and rate of clearance (CL) of all compounds.

The toxicity risk calculator locates fragments within the structure of the molecule that shows a potential toxicity risk. Toxicity risk parameters such as hERG K+-channel blocker, Human Hepatotoxicity (H-HT), Ames Mutagenicity (AMES), Skin sensitization and Drug-Induced Liver Injury (DILI) are computed for all the compounds. then the compounds having number 3,4,5,13,14,15,27,28, and 29 shows low hERG K+-channel blocking activity, all compound except 27 and 28 shows mild hepatotoxicity in high dose. Ames mutagenicity prediction result shows that except

compound number 23,24,27,28,29 and 32–50 showing mutagenicity and induces revertant colony growth. Skin sensitization prediction shows compounds other than 12,13,18,19,20,25,26 and 31 are skin nonsensitizer. Overall compounds are predicted to have mild toxicity risk levels. LD50 of acute toxicity predicted results for all compounds except compound number 23,41,42,43,45,46,47,48 and 50 are within the permissible limits (> 500 mg/kg) showing lower toxicity whereas above mentioned 9 compounds having LD50 value in between 51 and 500 mg/kg comes under toxicity level. The predicted drug likeliness and optimum synthetic accessibility score for all the compounds

a

b
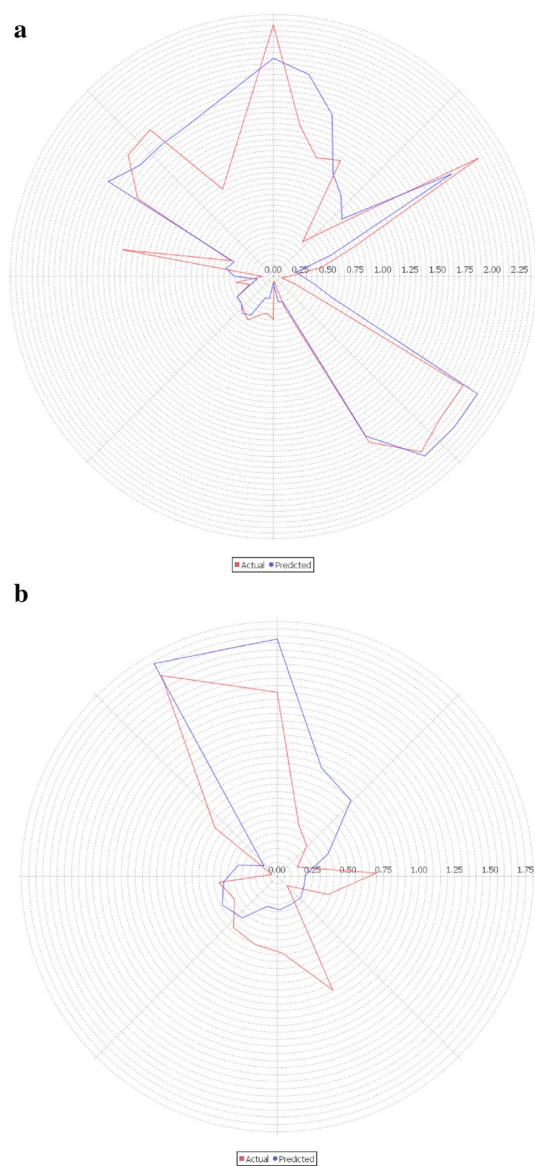
Fig. 8 **a** and **b** Actual Vs Predicted activity of Training and Test set for 3D-QSAR developed by SW-kNN MFA method

a

b
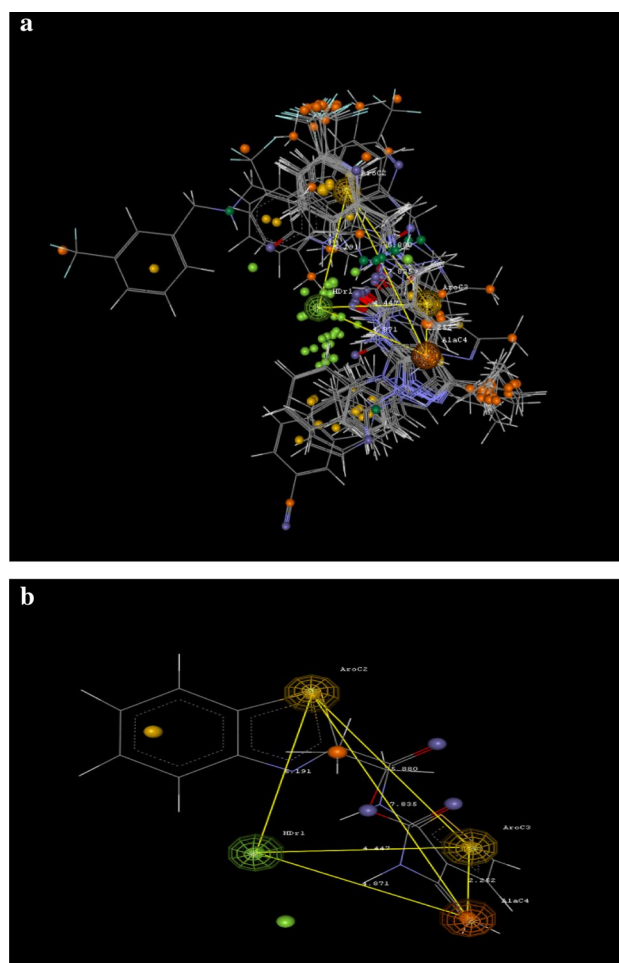
Fig. 9 **a** Pharmacophore hypothesis. **b** Distance based Pharmacophore identification

Table 7 Result of pharmacophore identification study

S. no	Pharmacophore features	Distance (A^0)
1	AroC2- AroC3	5.880
2	AroC3- AlaC4	2.258
3	AlaC4-Hdr1	4.871
4	Hdr1- AroC2	6.191
5	Hdr1- AroC3	4.447
6	AroC2- AlaC4	7.835

Table 8 Docking score of compounds

Compound no.	Grid dock score	Compound no	Grid dock score
1	−0.133192	26	−0.256474
2	0.469027	27	−3.138981
3	−0.568537	28	−2.005189
4	−0.680504	29	0.345981
5	−0.67013	30	1.16189
6	−0.941941	31	−0.410788
7	1.368081	32	1.255918
8	−2.922597	33	−1.077476
9	0.784597	34	1.58963
10	0.50095	35	−3.117195
11	0.647408	36	−0.460386
12	−1.185639	37	1.407166
13	−0.943537	38	1.907541
14	−0.90536	39	1.907541
15	−2.922597	40	−1.934221
16	−2.922597	41	2.009
17	0.353767	42	1.098156
18	−0.024999	43	1.6857
19	0.27378	44	2.540092
20	1.398964	45	2.398386
21	−0.975471	46	2.158539
22	−0.873857	47	1.006403
23	−1.334372	48	−1.281511
24	−0.000732	49	0.595667
25	−1.442838	50	−1.677993

suggest good druggability and easier synthesis of these compounds.

4 Conclusion

The combined computational approach is applied to give insight into the structural basis and inhibition mechanism for the series of compounds as DprE1 inhibitors antitubercular agents. Statistically significant QSAR models for both 2D and 3D QSAR provide a structural framework for understanding the relationship of chemical structure with the activity and exhibited a good correlation, predictive ability and satisfactory agreement between experimental and predicted activity of the training and test set molecules. The validated 2D-QSAR model was used to optimize the estate contribution, hydrophobicity, electrostatic and alignment independent requirements around the moiety to increase activity whereas 3D-QSAR model suggest that substitution of electronegative, electropositive and less bulky groups in particular site is preferable for antitubercular activity. Presence of two aromatic rings, one aliphatic and one hydrogen bond donor groups are the key pharmacophoric features for inhibition of DprE1

Fig. 10   a–e Binding model of compounds 8,15,16,27 and 35 with DprE1 target cavity
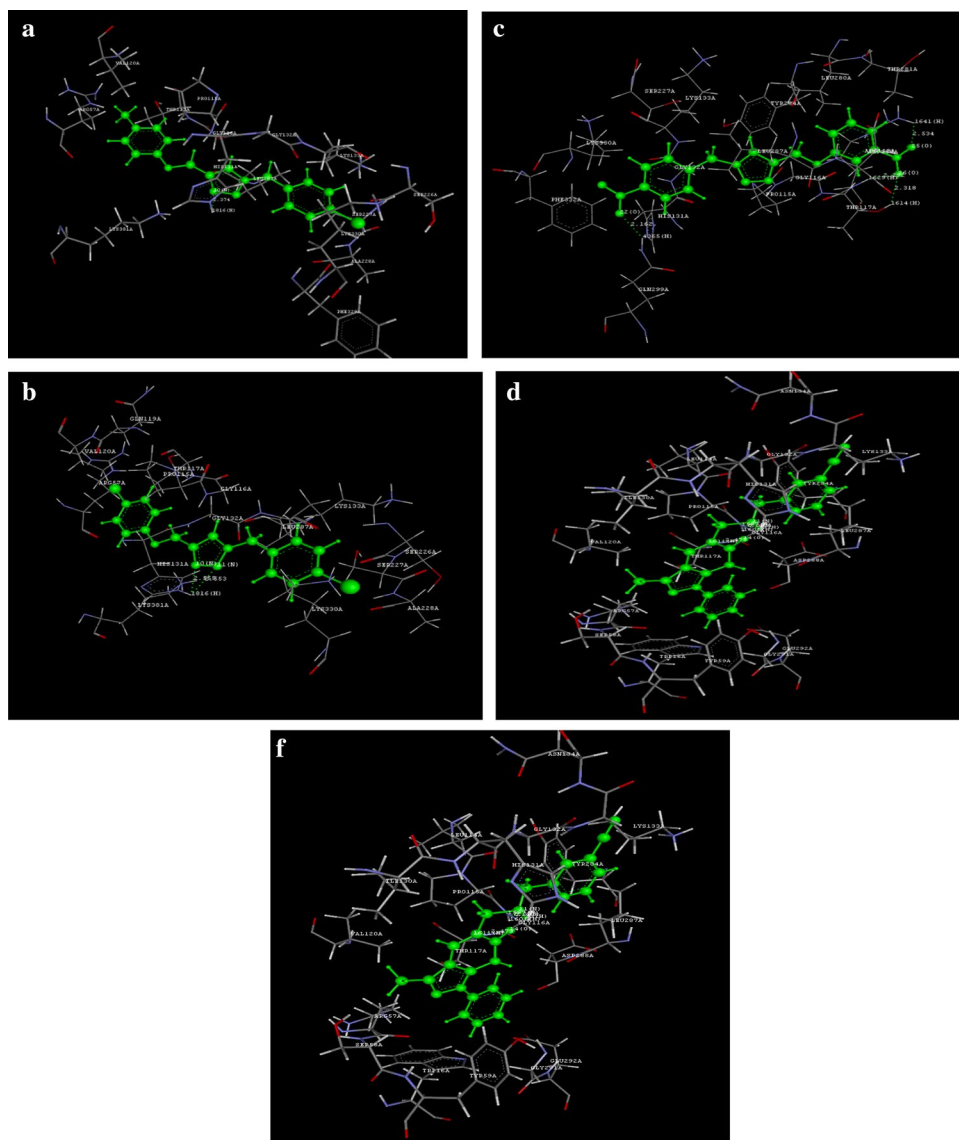
enzyme. Molecular docking study result shows five compounds of number 8,15,16,27 and 35 have significant interaction with the amino acid residues like Gly-116, His-131, Arg-118, Thr-117 and Gln-299 present at the active site of the target enzyme by forming non polar interaction (H-bond) suggest presence of H-bond forming atoms required for interactions between the ligands and the peptide residue. In silico prediction of drug likeliness and ADME-T risk profiling were within their acceptable limit confirm good druggability of these compounds and showing mild toxicity risk in high dose. The present computation approach will help to design new DprE1 inhibitors
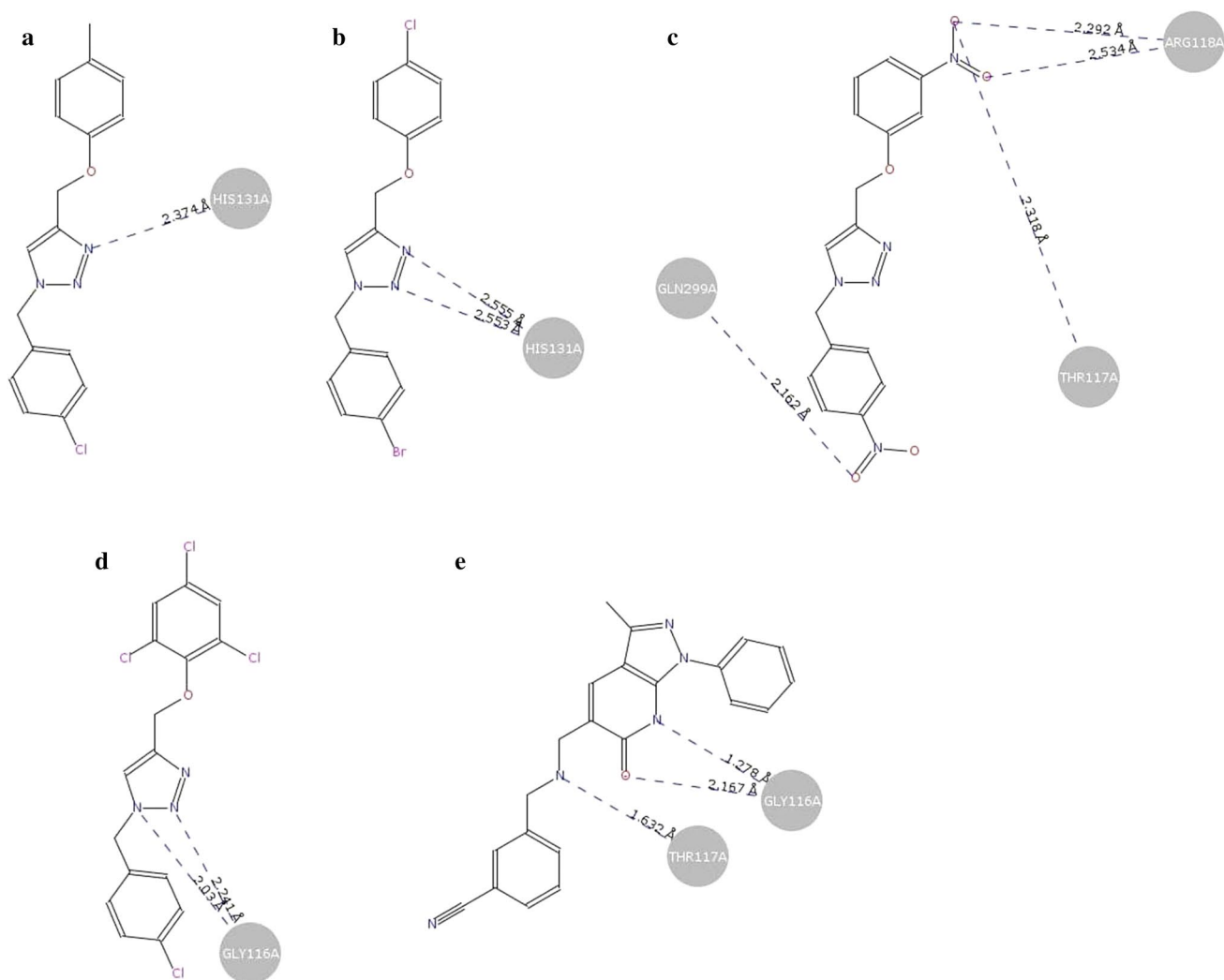
Fig. 11 **a–e** 2- Dimensional ligand interaction plot represents interaction of ligands (8,15,16,27 and 35) with different amino acid residues present on active site of DprE1 enzyme

Table 9 Ligand- target interaction result

Compound no	Residue atom	Ligand atom	Interaction type	Distance (A^0)
8	HIS131A	10-N	H-bond	2.373640
15	HIS131A	10-N	H-bond	2.555498
	HIS131A	11-N	H-bond	2.552875
16	THR117A	26-O	H-bond	2.318133
	ARG118A	26-O	H-bond	2.291650
	ARG118A	25-O	H-bond	2.533988
	GLN299A	22-O	H-bond	2.161704
27	GLY116A	11-N	H-bond	2.240558
	GLY116A	12-N	H-bond	2.030128
35	GLY116A	3-N	H-bond	1.278127
	GLY116A	14-O	H-bond	2.166816
	THR117A	11-N	H-bond	1.631831

Table 10 Pharmacokinetic features (Insilico ADME) prediction results

Com- pound no	Physiochemical properties			Absorption				Distribution			Metabolism										Eliminaion	
											Prediction of enzyme inhibition					Prediction as enzyme's substrate						
	LogS (log mol/ltr)	LogD$_{7.4}$	LogP	Papp (Caco-2 Permeability) in cm/sec	HIA	F	F'	PPB in %	VD (L/Kg)	BBB	CYP 1A2	CYP 3A4	CYP 2C9	CYP 2C19	CYP 2D6	CYP 1A2	CYP 3A4	CYP 2C9	CYP 2C19	CYP 2D6	T$_{1/2}$ (Hr)	CL mL/min/kg
1	−3.822	1.006	2.814	−4.653	+	+	+	88.19	−0.48	0.95	+	−	+	+	−	−	−	−	+	−	1.287	1.146
2	−3.788	0.989	2.814	−4.634	+	+	+	88.40	−0.49	0.94	+	−	+	+	−	−	−	+	−	−	1.263	1.177
3	−4.54	1.523	3.559	−4.585	+	+	+	91.99	−0.24	0.98	+	−	+	+	+	+	+	−	+	−	1.781	0.918
4	−4.499	1.511	3.559	−4.573	+	+	+	92.08	−0.24	0.98	+	−	−	+	+	+	−	−	+	−	1.796	0.9
5	−4.766	1.726	3.668	−4.598	+	+	+	92.36	−0.22	0.95	+	−	−	+	−	+	−	−	+	−	1.767	0.823
6	−4.228	1.179	3.122	−4.64	+	+	+	92.21	−0.43	0.87	−	+	+	+	−	,	,	+	+	+	1.495	1.27
7	−4.196	1.183	3.122	−4.621	+	+	+	92.75	−0.41	0.88	+	+	+	+	−	−	−	+	+	+	1.386	1.261
8	−4.8	1.554	3.867	−4.582	+	+	+	93.37	−0.18	0.97	+	+	+	+	+	+	+	+	+	−	1.874	0.911
9	−4.757	1.561	3.867	−4.559	+	+	+	93.59	−0.18	0.97	+	+	+	+	+	+	+	+	+	+	1.816	0.899
10	−4.883	1.772	3.976	−4.632	+	+	+	93.61	−0.13	0.91	+	+	+	+	+	+	+	+	+	+	1.884	0.822
11	−4.999	1.215	3.467	−4.683	+	+	+	92.46	−0.70	0.94	+	−	+	+	−	+	+	−	+	−	1.302	0.871
12	−4.969	1.213	3.467	−4.653	+	+	+	92.98	−0.70	0.95	+	+	+	+	+	+	+	−	+	−	1.289	0.859
13	−5.071	1.631	4.212	−4.573	+	+	+	93.83	−0.21	0.98	+	−	−	+	−	+	+	−	+	−	1.821	0.826
14	−5.011	1.635	4.212	−4.545	+	+	+	93.97	−0.21	0.984	+	−	+	+	+	+	+	−	+	−	1.825	0.835
15	−5.289	1.848	4.321	−4.627	+	+	+	93.91	−0.17	0.96	+	−	+	+	−	+	+	+	+	−	1.741	0.706
16	−4.215	0.951	2.722	−4.618	+	+	+	87.44	−0.95	0.943	+	−	+	+	−	−	−	−	−	−	1.02	0.892
17	−4.166	0.94	2.722	−4.605	+	+	+	87.39	−0.96	0.943	+	−	+	+	+	−	−	−	−	−	0.978	0.92
18	−4.962	1.238	3.467	−4.65	+	+	+	93.29	−0.72	0.955	+	+	+	+	+	−	−	−	−	−	1.209	0.96
19	−4.948	1.23	3.467	−4.648	+	−	+	93.26	−0.72	0.955	+	−	+	+	+	−	−	−	−	−	1.233	0.963
20	−4.867	1.348	3.576	−4.661	+	+	+	93.18	−0.78	0.903	+	+	+	+	−	−	−	−	−	−	1.293	0.785
21	−5.294	1.237	3.775	−4.663	+	+	+	93.43	−0.66	0.917	+	+	+	+	−	+	+	+	+	+	1.335	0.946
22	−5.286	1.234	3.775	−4.639	+	+	+	93.45	−0.65	0.929	+	+	+	+	+	+	+	+	+	+	1.342	0.916
23	−5.403	1.724	4.521	−4.62	+	+	+	94.05	−0.17	0.978	+	+	+	+	+	+	+	+	+	+	1.897	0.887
24	−5.489	1.911	4.63	−4.648	+	+	+	93.93	−0.18	0.954	+	+	+	+	+	+	+	+	−	−	1.847	0.763
25	−5.285	1.603	4.774	−4.689	+	+	+	92.76	−0.78	0.894	+	−	+	+	−	−	−	−	−	−	1.418	0.575
26	−5.284	1.6	4.774	−4.665	+	+	+	92.77	−0.79	0.912	+	+	+	+	+	+	−	−	+	−	1.435	0.594
27	−6.059	1.906	5.519	−4.633	+	+	+	93.67	−0.31	0.955	+	−	+	−	−	+	−	−	+	−	1.644	0.696
28	−6.047	1.901	5.519	−4.623	+	+	+	93.65	−0.31	0.963	+	−	+	−	−	+	−	+	+	−	1.637	0.685
29	−6.212	1.963	5.628	−4.666	+	+	+	93.91	−0.37	0.921	+	+	+	+	+	−	−	+	+	−	1.593	0.65
30	−4.793	1.326	3.931	−4.667	+	+	+	90.82	−1.17	0.651	+	−	−	−	−	−	−	+	−	−	1.429	0.079
31	−5.299	1.582	4.785	−4.695	+	+	+	92.88	−0.86	0.565	+	+	+	+	+	−	−	−	−	+	1.587	0.49

Table 10 (continued)

Com-pound no	Physiochemical properties			Absorption				Distribution			Metabolism											Eliminaion	
											Prediction of enzyme inhibition					Prediction as enzyme's substrate							
	LogS (log-mol/ltr)	LogD$_{7.4}$	LogP	Papp (Caco-2 Permeability) in cm/sec	HIA	F	F'	PPB in %	VD (L/Kg)	BBB	CYP 1A2	CYP 3A4	CYP 2C9	CYP 2C19	CYP 2D6	CYP 1A2	CYP 3A4	CYP 2C9	CYP 2C19	CYP 2D6	T$_{1/2}$ (Hr)	CL mL/min/kg
32	−4.81	2.868	3.447	−5.015	+	+	+	95.42	0.314	0.977	+	+	+	+	+	+	−	+	+	−	1.942	2.05
33	−4.015	2.321	2.83	−4.985	+	+	+	92.26	0.196	0.982	+	−	−	+	−	+	−	+	+	−	1.812	2.11
34	−4.287	2.794	3.139	−4.989	+	+	+	94.87	0.237	0.971	+	+	+	+	−	+	−	+	+	−	1.815	2.099
35	−4.044	2.093	2.702	−5.079	+	+	+	91.58	0.086	0.924	+	+	+	+	−	+	+	−	+	−	1.739	2.005
36	−4.213	2.414	2.969	−5.032	+	+	+	94.32	−0.06	0.99	+	+	+	+	−	+	−	−	+	−	1.791	1.965
37	−5.149	2.749	3.849	−5.11	+	+	+	96.27	−0.28	0.97	+	+	+	+	−	+	+	−	+	−	1.786	1.671
38	−5.31	2.799	4.157	−5.121	+	+	+	96.42	−0.27	0.965	+	+	+	+	−	+	+	−	+	−	1.794	1.689
39	−4.515	2.605	3.244	−5.075	+	+	+	95.24	−0.22	0.963	+	+	−	+	−	+	+	−	+	−	1.695	1.619
40	−4.218	1.977	2.397	−4.521	+	+	+	91.25	−0.23	0.979	+	+	−	+	−	+	−	−	+	−	1.629	1.708
41	−5.576	2.943	4.418	−5.195	+	+	+	96.07	−0.18	0.931	+	+	+	+	+	+	+	−	+	−	1.745	1.677
42	−4.971	2.855	3.813	−5.101	+	+	+	95.89	−0.15	0.942	−	+	+	+	−	+	+	−	+	−	1.683	1.521
43	−5.49	2.869	4.557	−5.095	+	+	+	94.48	−0.21	0.907	−	+	+	+	+	+	+	−	+	−	1.84	1.599
44	−5.507	2.962	4.557	−5.096	+	+	+	95.98	−0.22	0.953	−	+	+	+	−	+	+	−	+	−	1.867	1.607
45	−5.148	2.85	4.121	−5.125	+	+	+	96.24	−0.10	0.803	−	+	+	+	−	−	+	−	+	−	1.679	1.536
46	−5.532	2.946	4.726	−5.174	+	+	+	95.65	−0.16	0.9	+	+	+	+	−	+	+	−	+	−	1.79	1.678
47	−5.315	2.73	4.157	−5.109	+	+	+	96.08	−0.26	0.974	+	+	+	+	−	+	+	+	+	−	1.802	1.729
48	−6.51	2.992	4.972	−5.149	+	+	+	96.84	−0.26	0.929	+	+	+	+	−	+	+	−	+	−	1.81	1.774
49	−5.149	2.749	3.849	−5.11	+	+	+	96.27	−0.28	0.977	+	+	+	+	−	+	+	−	+	−	1.786	1.671
50	−5.0	2.767	4.117	−4.709	+	+	+	95.37	0.14	0.972	−	−	−	+	+	+	+	−	+	+	1.753	1.481

Log^S=Solubility (optimum >−4 log mol/l), LogD$_{7,4}$=Distribution coefficient (<1 Solubility high), LogP=Distribution coefficient (0 <LogP <5), Papp (Caco-2Permeability)=Human intestinal absorption (optimal range:higher than −4.50 Log unit), HIA=Human Intestinal Absorption (≥ 30%:HIA is +ve; < 30%: HIA is −ve), F = 20%Bioavailability (≥ 20%: F20 is +ve; < 20%: F20 is −ve), F'= 30% Bioavailability (≥30%: F'30 is+ve; < 30%: F'30 is −ve), PPB=Plasma Protein Binding, VD=Volume of Distribution (optimal range 0.04–20L/kg), BBB=Blood–Brain Barrier (BB ratio ≥ 0.1:BBB is +ve; BB ratio ≤ 0.1: BBB is −ve), T$_{1/2}$=Half life time (>8 h: high; 3 h <Cl <8 h: moderate; <3 h:low), CL=Clearance rate (>15 mL/min/kg: high; 5 mL/min/kg <Cl <15 mL/min/kg: moderate; < 5 mL/min/kg: low)

Table 11 Predicted toxicity parameters and Lipinski's rule of five drug likeliness of compounds

Compound no	Toxicity parameters						Drug likeliness							
	hERG blockers	H-HT	AMES	SkinSen	LD50 (mg/kg)	DILI	Molecular weight (<500 amu)	HB acceptor (<10)	HB donor (<05)	No. of rotatable bonds (0–15)	TPSA (<140Å²)	Drug likeliness	Lead likeliness	Synthetic accessibilit score
1	+	+	+	−	1617.335	+	310.313	5	0	6	85.76	Yes	Yes	2.62
2	+	+	+	−	1666.479	+	310.313	5	0	6	85.76	Yes	Yes	2.67
3	−	+	+	−	1166.206	+	299.761	3	0	5	39.94	Yes	Yes	2.55
4	−	+	+	−	1163.523	+	299.761	3	0	5	39.94	Yes	Yes	2.58
5	−	+	+	−	972.358	+	344.212	3	0	5	39.94	Yes	Yes	2.58
6	+	+	+	−	1803.02	+	324.34	5	0	6	85.76	Yes	Yes	2.72
7	+	+	+	−	1753.883	+	324.34	5	0	6	85.76	Yes	Yes	2.77
8	+	+	+	−	1187.507	+	313.788	3	0	5	39.94	Yes	Yes	2.65
9	+	+	+	−	1234.913	+	313.788	3	0	5	39.94	Yes	Yes	2.69
10	+	+	+	−	942.265	+	358.239	3	0	5	39.94	Yes	Yes	2.69
11	+	+	+	+	1391.6	+	344.758	5	0	6	85.76	Yes	Yes	2.63
12	+	+	+	+	1547.086	+	344.758	5	0	6	85.76	Yes	Yes	2.68
13	−	+	+	−	1011.613	+	334.206	3	0	5	39.94	Yes	Yes	2.57
14	−	+	+	−	1011.613	+	334.206	3	0	5	39.94	Yes	No	2.60
15	−	+	+	−	944.596	+	378.657	3	0	5	39.94	Yes	No	2.60
16	+	+	+	−	1554.56	+	355.31	7	0	7	131.58	Yes	No	2.85
17	+	+	+	−	1481.179	+	355.31	7	0	7	131.58	Yes	No	2.90
18	+	+	+	+	1450.489	+	344.758	5	0	6	85.76	Yes	Yes	2.78
19	+	+	+	+	1440.504	+	344.758	5	0	6	85.76	Yes	Yes	2.80
20	+	+	+	−	1125.079	+	389.209	5	0	6	85.76	Yes	No	2.80
21	+	+	+	−	1235.476	+	358.785	5	0	6	85.76	Yes	No	2.75
22	+	+	−	−	1284.796	+	358.785	5	0	6	85.76	Yes	No	2.80
23	+	+	−	−	348.233	+	348.233	3	0	5	39.94	Yes	No	2.69
24	+	+	+	−	847.312	+	392.684	3	0	5	39.94	Yes	No	2.71
25	+	+	+	+	945.433	+	413.648	5	0	6	85.76	Yes	No	2.75
26	+	+	+	+	921.788	+	413.648	5	0	6	85.76	Yes	No	2.79
27	−	−	−	−	676.719	+	403.096	3	0	5	39.94	Yes	No	2.70
28	−	−	−	−	614.34	+	403.096	3	0	5	39.94	Yes	No	2.73
29	−	+	−	−	683.658	+	447.547	3	0	5	39.94	Yes	No	2.72
30	+	+	+	−	1208.542	+	607.102	7	0	7	131.58	Yes	No	2.96
31	+	+	+	+	1461.702	+	641.001	5	0	6	85.76	Yes	No	2.91
32	+	+	−	−	855.251	+	372.472	3	2	5	62.71	Yes	No	3.13
33	+	+	−	−	758.728	+	344.418	3	2	5	62.71	Yes	Yes	2.89
34	+	+	−	−	789.628	+	358.445	3	2	5	62.71	Yes	No	3.02

Table 11 (continued)

Compound no	Toxicity parameters						Drug likeliness							Drug likeliness	Lead likeliness	Synthetic accessibilit score
	hERG blockers	H-HT	AMES	SkinSen	LD50 (mg/kg)	DILI	Molecular weight (<500 amu)	HB acceptor (<10)	HB donor (<05)	No. of rotatable bonds (0–15)	TPSA (<140Å²)					
35	+	+	−	−	784.386	+	369.428	4	2	5	86.50			Yes	No	3.01
36	+	+	−	−	679.513	+	362.408	4	2	5	62.71			Yes	No	2.92
37	+	+	−	−	502.73	+	412.415	6	2	6	62.71			Yes	No	3.08
38	+	+	−	−	430.388	+	426.442	6	2	6	62.71			Yes	No	3.20
39	+	+	−	−	731.768	+	413.403	7	2	6	75.60			Yes	No	3.23
40	+	+	−	−	1008.079	+	350.344	5	2	6	62.71			Yes	No	2.72
41	+	+	−	−	398.037	+	438.453	6	2	7	62.71			Yes	No	3.31
42	+	+	−	−	241.491	+	439.441	7	2	7	75.60			Yes	No	3.29
43	+	+	−	−	240.097	+	456.443	7	2	7	62.71			Yes	No	3.37
44	+	+	−	−	175.544	+	456.443	7	2	7	62.71			Yes	No	3.35
45	+	+	−	−	183.885	+	453.468	7	2	7	75.60			Yes	No	3.42
46	+	+	−	−	260.975	+	452.48	6	2	7	62.71			Yes	No	3.45
47	+	+	−	−	333.319	+	426.442	6	2	7	62.71			Yes	No	3.22
48	+	+	−	−	183.455	+	454.496	6	2	7	62.71			Yes	No	3.50
49	+	+	−	−	502.73	+	412.415	6	2	7	62.71			Yes	No	3.08
50	+	+	−	−	249.415	+	438.453	6	1	4	53.92			Yes	No	3.28

hERG Blockers = the human Ether-a-go-go-related Gene blocker (hERG Blockers (+) = blocker. hERG Blockers (−) = non blocker, H-HT = Human Hepatotoxicity (H-HT (−) = nontoxic, H-HT (+) = toxic), AMES = Ames Mutagenicity (Ames (+) = induces revertant colony growth & Ames (−) = no growth), SkinSen = Skin sensitization {(+) = Sensitizer & (−) = Non sensitizer}, LD50 = LD50 of acute toxicity (High-toxicity: 1 ~ 50 mg/kg;Toxicity: 51 ~ 500 mg/kg;low-toxicity: 501 ~ 5000 mg/kg), DILI = Drug induced liver injury, HB Acceptor = Hydrogen bond acceptor, HB Donor = Hydrogen bond donor, TPSA = Topological polar surface area, Synthetic accessibility Score (1–10 (veryeasy), > 10 (verydifficult)

based on the results of QSAR studies. Thus, these compounds have rationalized the possible structural requirement for better binding interactions with target site and need further lead optimization for designing of more potent DprE1 inhibitors.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

References

1. Villemagne B, Crauste C, Flipo M, Baulard AR, Deprez B, Willand N (2012) Tuberculosis the drug development pipeline at a glance. Eur J Med Chem 51:1–16
2. Bansal R, Sharma D, Singh R (2018) Tuberculosis and its treatment: an overview. Mini Rev Med Chem 18(1):58–71
3. Fogel N (2015) Tuberculosis: a disease without boundaries. Tuberculosis 95(5):527–531
4. WHO Global Tuberculosis Report (2018). https://www.who.int/tb/publications/global_report/en/.
5. Tuberculosis. WHO (2010).https://www.who.int/tb/publications/2010/factsheet_tb_2010.pdf.
6. Balganesh TS, Alzari PM, Cole ST (2008) Rising standards for tuberculosis drug development. Trends Pharmacol Sci 29:576–581
7. Brennan PJ (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis 83(1–3):91–97
8. Kumar K, Kon OM (2017) Diagnosis and treatment of tuberculosis: latest developments and future priorities. Ann Res Hosp 1(37):1–15
9. Orme IM (2011) Development of new vaccines and drugs for TB: limitations and potential strategic errors. Further Microbiol 6:161–177
10. Lemke TL, Williams DA, Roche VF, Zito SW (2008) Anti mycobacterial agents. In: Foye's principles of medicinal chemistry, 6th edn. L. Williams and Wilkins, a Wolters Kluwer Business, Phildelphia, pp 1127–1146
11. Tripathi KD (2013) Anti tuberculosis. In: Essentials of medical pharmacology, 7th edn. Jaypee Brothers Medical Publishers Pvt. Ltd, India, pp 765–770
12. Vaghela JF, Kapoor SK, Kumar A, TewariDass R, Khanna A, Bhatnagar AK (2015) Home based care to multi-drug resistant tuberculosis patients: a pilot study. Indian J Tuberc 62(2):91–96
13. Patel SV, Nimavat KB, Patel BA, Shukla LK, Shringarpure KS, Mehta KG, Joshi CC (2016) Treatment outcome among cases of multidrug-resistant tuberculosis (MDR TB) in Western India: a prospective study. J Infect Public Health 9(4):478–484
14. Mikusova K, Huang H, Yagi TH, Vereecke D, D'Haeze W, Scherman MS, Brennan PJ, McNeil MR, Crick DC (2005) Decaprenylphosphoryl Arabinofuranose, the donor of the d- Arabinofuranosyl residues of *Mycobacterial arabinan*, is formed via a two-step epimerization of decaprenylphosphoryl ribose. J Bacteriol 187:8020–8025
15. Manina G, Pasca MR, Buroni S, De Rossi E, Riccardi G (2010) Decaprenylphosphoryl-β-D-ribose 2′-epimerase from *Mycobacterium tuberculosis* is a magic drug target. Curr Med Chem 17:3099–3108
16. Incandela ML, Perrin E, Fondi M, de Jesus LRAL, Mori G, Moiana A, Gramegna M, Fani R, Riccardi G, Pasca MR (2013) DprE1, a new taxonomic marker in mycobacteria. FEMS Microbiol Lett 348:66–73
17. Riccardi G, Pasca MR, Chiarelli LR, Manina G, Mattevi A, Binda C (2013) The DprE1 enzyme, one of the most vulnerable targets of *Mycobacterium tuberculosis*. Appl Microbiol Biotechnol 97:8841–8848
18. Miroslav B, Ivana C, Mukherjee R, Gaelle SK, Stanislav H, Adela B, Emoke K, Veronika M, Zuzana S, Michal S, Joao N, Jana K, Cole ST, Mikusova K (2015) DprE1 is a vulnerable tuberculosis drug target due to its cell wall localization. ACS Chem Biol 10:1631–1636
19. Trefzer C, Skovierova H, Buron S, Bobovska A, Nenci S, Molteni E, Pojer F, Pasca MR, Makarov V, Cole ST, Riccardi G, Mikusova K, Johnsson K (2012) Benzothiazinones are suicide inhibitors of mycobacterial decaprenylphosphoryl-β-d-ribofuranose 2′-Oxidase DprE. J Am Chem Soc 134:912–915
20. Batt SM, Jabeen T, Bhowruth V, Quill L, Lund PA, Eggeling L, Alderwick LJ, Fütterer K, Besra GS (2012) Structural basis of inhibition of *Mycobacterium tuberculosis* DprE1 by benzothiazinone inhibitors. Proc Natl Acad Sci 109:11354–11359
21. Chikhale R, Menghani S, Babu R, Bansode R, Bhargavi G, Karodia N, Rajasekharan MV, Paradkar A, Khedekar PO (2015) Development of selective DprE1 inhibitors: design, synthesis, crystal structure and antitubercular activity of benzothiazolylpyrimidine-5- carboxamide. Eur J Med Chem 96:30–46
22. Correa-Basurto J, Bello M, Rosales-Hernandez MC, Hern_andez-Rodríguez M, Nicol_as-V_azquez I, Rojo-Domínguez A, Flores-Sandoval CA (2014) QSAR, docking, dynamic simulation and quantum mechanics studies to explore the recognition properties of cholinesterase binding sites. Chem. Biol. Interact 209:1–13
23. Wang Z, Cheng L, Kai Z, Wu F, Liu Z, Cai M (2014) Molecular modeling studies of atorvastatin analogues as HMGR inhibitors using 3D-QSAR, molecular docking and moleculardynamics simulations. Bioorg Med Chem Lett 24:3869–3876
24. Singh A, Goyal S, Jamal S, Subramani B, Das M, Admane N, Grover A (2016) Computational identification of novel piperidine derivatives as potential HDM2 inhibitors designed by fragment- based QSAR, molecular docking and molecular dynamics simulations. Struct Chem 27:993–1003
25. Sharma MC, Sharma S (2016) Molecular modeling studies of 3-acyl-2-phenylamino-1,4- dihydroquinolin-4-one derivatives as phosphatase SerB653 inhibitors. Med Chem Res 25:2119–2126
26. Sharma MC (2016) Computational design of novel renin inhibitors of indole-3- Carboxamide derivatives through QSAR studies. Netw Model Anal Health Inform Bioinforma 5:1–12
27. Sharma MC (2014) Identification of 3-Nitro-2,4,6-trihydroxy-benzamide derivatives as photosynthetic electron transport

inhibitors by qsar and pharmacophore studies. Interdiscip Sci Comput Life Sci 6:1–13

28. Huang LL, Han J, Ran JX, Chen XP, Wang ZH, Wu FH (2018) 3D-QSAR, molecular docking and molecular dynamics simulations of oxazepane amidoacetonitrile derivatives as novel DPPI inhibitors. J Mol Struct 1168:223–233

29. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 9:1639–1662

30. Golbraikh A, Tropsha A (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des 16:357–369

31. Hadjipavlou-Litina D (1998) Review, revaluation, and new results in quantitative structure– activity studies of anticonvulsants. Med Res Rev 18:91–119

32. Shaikh MH, Dnyaneshwar DS, Nawale L, Sarkar D, Khan FAK, Sangshetti JN, Shingate BB (2015) 1,2,3- Triazole derivatives as anti tubercular agents: synthesis, biological evaluation and molecular docking study. Med Chem Comm 6:1–34

33. Panda M, Ramachandra S, Ramchandran V, Shirude PS, Humnabadakar V, Nagalapur K, Sharma S, Kaur P, Guptha S, Narayan A, Mahadevaswamy J, Ambady A, Hegde A, Rudrapatna SS, Hosagrahara VP, Sambandamurthy VK, Raichukar A (2014) Discovery of pyrazolopyridones as a novel class of noncovalent DprE1 inhibitor with potent anti- mycobacterial activity. J Med Chem 57:4761–4771

34. VLife MDS 4.6 Molecular design suite (2018). Vlife Sciences Technologies Pvt. Ltd. Pune, India

35. Halgren TA (1996) Merck molecular force field. III. Molecular geometries and vibrational frequencies. J Comp Chem 17:553–586

36. Holland JH (1992) Genetic algorithms. Sci Am 267:66–72

37. Dong H, Liu J, Liu X, Yu Y, Cao S (2017) Molecular docking and QSAR analyses of aromatic heterocycle thiosemicarbazone analogues for finding novel tyrosinase inhibitors. Bioorg Chem 75:106–117

38. Tang HJ, Yang L, LiJH CJ (2016) Molecular modelling studies of 3, 5- dipyridyl-1, 2, 4- triazole derivatives as xanthine oxidoreductase inhibitors using 3D-QSAR, Topomer CoMFA, molecular docking and molecular dynamic simulations. J Taiwan Inst Chem Eng 68:64–73

39. Wang JL, Cheng LP, Wang TC, Deng W, Wu FH (2017) Molecular modelling study of CP- 690550 derivatives as JAK3 kinase

inhibitors through combined 3D QSAR, molecular docking, and dynamics simulation techniques. J Mol Graph Model 72:178–186

40. Dong H, Liu J, Liu X, Yu Y, Cao S (2018) Combining molecular docking and QSAR studies for modelling the anti-tyrosinase activity of aromatic heterocycle thiosemicarbazone analogues. J Mol Struct 74:304–326

41. Vepuri SB, Anbazhagan S, Naresh P, Divya D (2012) Pharmacophore modeling ad docking based QSAR studies of Aryl Amidino Isoxazoline derivatives to design potential FXa inhibitors. Am J Bio Res 2:11–20

42. Behera DK, Behera PM, Acharya L, Dixit A (2017) Pharmacophore modelling, virtual screening and molecular docking studies on PLD1 inhibitors. SAR QSAR Environ Res 28:991–1009

43. Khan MF, Verma G, Akhtar W, Shaquiquzzaman M, Akhter M, Rizvi MA, Alam MM (2016) Pharmacophore modelling, 3D-QSAR, docking study and ADME a prediction of acyl 1,3,4-thizole amides and sulphonamides as antitubulin agents. Arab. J. Chem. https://doi.org/10.1016/j.arabjc.2016.09.019

44. Safarizadeh H, Garkani-Nejad Z (2019) Molecular docking, molecular dynamics simulations and QSAR studies on some of 2-arylethenylquinoline derivatives for inhibition of Alzheimer's amyloid-beta aggregation: insight into mechanism of interactions and parameters for design of new inhibitors. J Mol Graph Model 87:129–143

45. Dong J, Wang NN, Yao ZJ, Zhang L, Cheng Y, Ouyang D, Lu AP, Cao DS (2018) ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. J Cheminform 10:1–11

46. Yadav DK, Saloni SK, Singh H, Kim M, Sharma P, Misra S, Khan F (2017) Molecular docking, QSAR and ADMET studies of Withnolide analogs against brest cancer. Drug Des Dev Ther 11:1859–1870

47. Athar M, Lone MY, Khedkar VM, Jha PC (2016) Pharmacophore model prediction, 3D-QSAR and molecular docking studies on vinyl sulfones targeting Nrf2-mediated gene transcription intended for Anti-Parkinson drug design. J Biomol Struct Dyn 34:1282–1297

48. Zhang J, Shan Y, Pan X, Wang C, Xu W, He L (2011) Molecular docking, 3D-QSAR studies, and in silico ADME prediction of p-aminosalicylic acid derivatives as neuraminidase inhibitors. Chem Biol Drug Des 78:709–717