



Multiple 3D- and 2D-quantitative structure–activity relationship models (QSAR), theoretical study and molecular modeling to identify structural requirements of imidazopyridine analogues as anti-infective agents against tuberculosis

Suraj N. Mali¹ · Anima Pandey¹ · Babu R. Thorat² · Chin-Hung Lai³

Received: 16 December 2021 / Accepted: 11 January 2022 / Published online: 24 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Tuberculosis (TB), an infectious remains a global health burden till date. Considering immense importance of theoretical tools in computer aided-drug designing, the current study focuses on common pharmacophore and QSAR analysis of 38 imidazopyridine analogues as anti-TB agents. Our developed atom-based, field-based, and multilinear regression based-QSAR models showed high values for statistical robustness for internal as well as external validations (a correlation coefficient: $R^2 > 0.9$, least standard deviations, higher Fischer coefficient, and cross-validation correlation coefficient: $Q^2 > 0.5$). From our ZINC-Drug-like analysis, we were retained with 5 hits (VS1-VS-5), among them VS-4 molecule was found to have high potency (predicted pIC_{50} (μM) value: 7.96 (against MTB H37Rv ATCC 27,294)) with good theoretical properties (high softness, and low hardness values). From our designed analogues (S1-S10), analogue S-10 was retained with high potency as well as good pharmacokinetics to act as good anti-mycobacterial agent in future.

Keywords Anti-infectives · Imidazopyridine · *Mycobacterium tuberculosis* · QSAR modelling · Theoretical study · DFT

Abbreviations

TB	Tuberculosis
QSAR model	Quantitative structure–activity relationship model
DFT	Density functional theory
MDR-TB	Multi-drug-resistant tuberculosis
XDR-TB	Extensively drug-resistant tuberculosis
3D-QSAR model	3D-quantitative structure–activity relationship models
CPH	Common pharmacophore hypothesis
R^2	Correlation coefficient
Q^2	Cross-validation correlation coefficient

RMSE	Root mean square error
ADMET	Absorption, distribution, metabolism, excretion and toxicity properties
MIC	Minimum inhibitory concentration
IC_{50}	Inhibitory concentration
PLS	Partial least squares

Introduction

Tuberculosis (TB), an infectious disease caused by *Mycobacterium*, remains a global health burden since many decades [1]. With the emergence of multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR TB) strains, these infections have been amplified further and became difficult to cure with the conventional anti-TB therapy [2]. Moreover, TB has an active association with acquired immunodeficiency syndrome (AIDS) [3]. Such scenario of co-infection was accounted for 26% of all AIDS-associated deaths [2]. The WHO (World Health Organization) statistics on TB for the year 2020 says that there were 1.5 million people died from tuberculosis (including 214,000 AIDS-related deaths). Current pharmacotherapy for

✉ Suraj N. Mali
mali.suraj1695@gmail.com

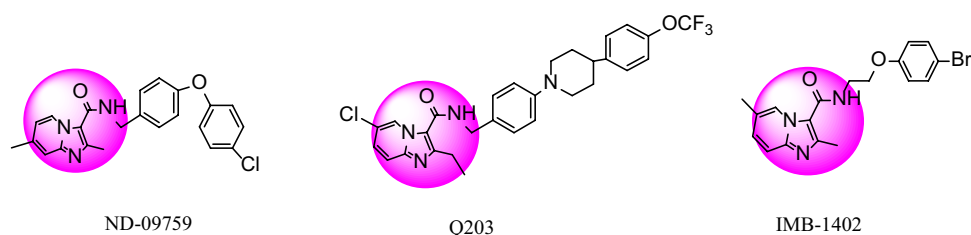
Anima Pandey
apandey@bitmesra.ac.in

¹ Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Mesra, India

² Department of Chemistry, Government College of Arts and Science, 431 001 Aurangabad (M.S.), India

³ Department of Medical Applied Chemistry, Chung Shan Medical University, Taichung 40241, Taiwan

Fig. 1 Structures of ND-09759, Q203, and IMB-1402



TB includes several drugs which has severe adverse side effects. Moreover, these drugs are now become ineffective due to resistant strains [4]. Thus, to deal with these issues, there is an urgent need to develop new chemical entities (NCEs) with unique mechanism of actions.

Imidazopyridine motifs are very crucial in the rational drug design and developments of newer drugs [5]. Recently, this scaffold has tremendously explored for central nervous system (CNS), antidiabetic, antitubercular, antiviral, anticancer agents, etc. [5]. An optimization of imidazopyridine heterocyclic system would serve as important medicinal chemistry tool to enhance further potency of drug moiety. Q203 and ND09759 were two clinical drug candidates, which showed strong activity against resistant strains of TB (Fig. 1) [1]. The IMB-1402 candidate also showed acceptable safety parameters (Fig. 1).

From several decades, computer-aided drug design (CADD) has been immensely explored and benefited for new drug discovery, structural optimization, and the target identification. Considering large number of supportive literatures [6–17], it is very clear that CADD approach facilitates an economical, less time-consuming, and successful outputs with the help of varieties of algorithms and ideas. Thus, this work focuses on incorporation of CADD approach in anti-TB drug discovery processes using ligand-based as well as structure-based drug design techniques. The QSAR approach would facilitate medicinal chemists to retrieve exact molecular characteristics required for biological activity and thus, would serve as significant tool in drug designing.

In our present study, we have carried out PHASE (Schrödinger, 2021) generated pharmacophore and multiple QSAR analysis (Figs. 2, 3, 4, 5, 6, 7; Tables 1, 2, 3, 4, 5, 6, 7) [17–20]. The generated pharmacophore model signifies key specific characteristics required for imidazopyridines to act as potent antimycobacterial inhibitors. Our current Atom as well as Field based 3D-QSAR models correlates key features required with inhibitory potencies of molecules. Thus, from these information, one can design more potent imidazopyridines. Moreover, with the help of designed 2D- and 3D-QSAR models, we come up with newer designed 10 (S1–S10) imidazopyridines analogues (Table 8) with better in silico pharmacokinetics and good predictive potencies. From the originally reported best hit (molecule 24), we screened ZINC drug like database and retrieved with top 5 imidazopyridines (Table 9). These hit molecules were then analyzed theoretically using DFT approach (Figs. 8, 9, 10). From theoretical properties, the best probable hit molecule VS-4 is also reported herein (Table 10).

Materials and methods

Softwares

In our current study, the developments of common pharmacophore hypothesis (CPH) and 3D-QSAR (quantitative structure–activity relationships (QSAR)) models were carried out using PHASE module (Schrödinger, LLC,

Fig. 2 (a) Pharmacophore model (HHPRR_1) generated by PHASE. and (b) The HHPRR_1 model illustrates hydrophobic feature (H4H5; green color), positive feature (P8; blue), and aromatic rings (R10R11, brown color) features. All active ligands overlapped on the generated model HHPRR

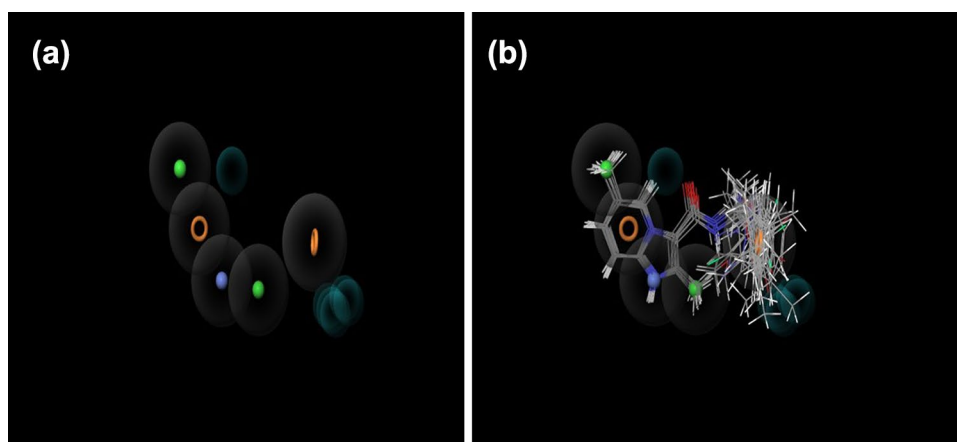


Fig. 3 Graphical presentation of actual versus predicted pIC_{50} of (a) training and (b) test set molecules for obtained atom-based 3D-QSAR model

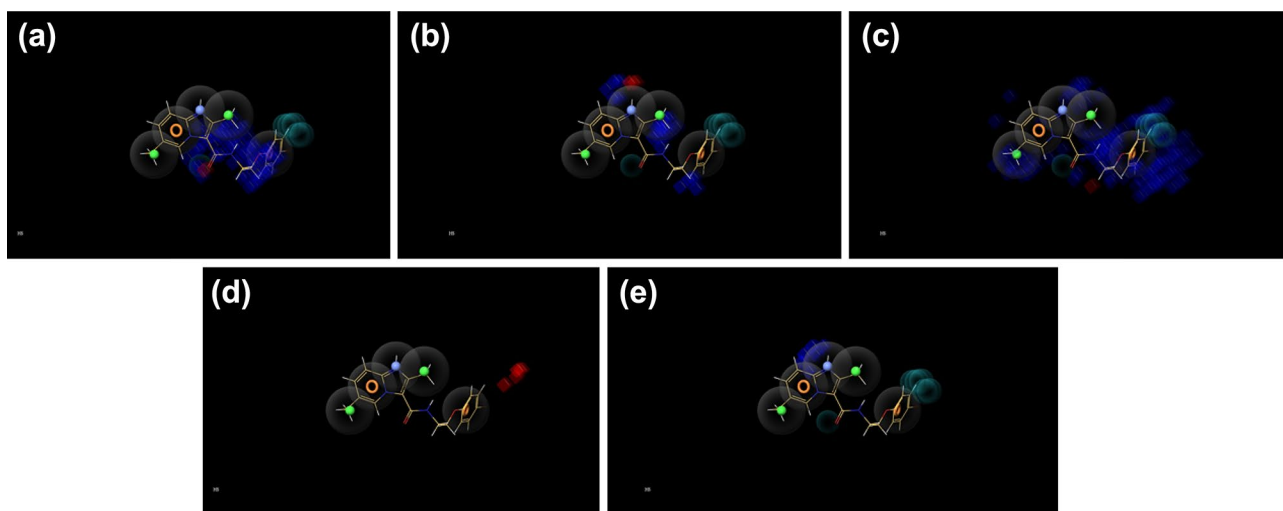
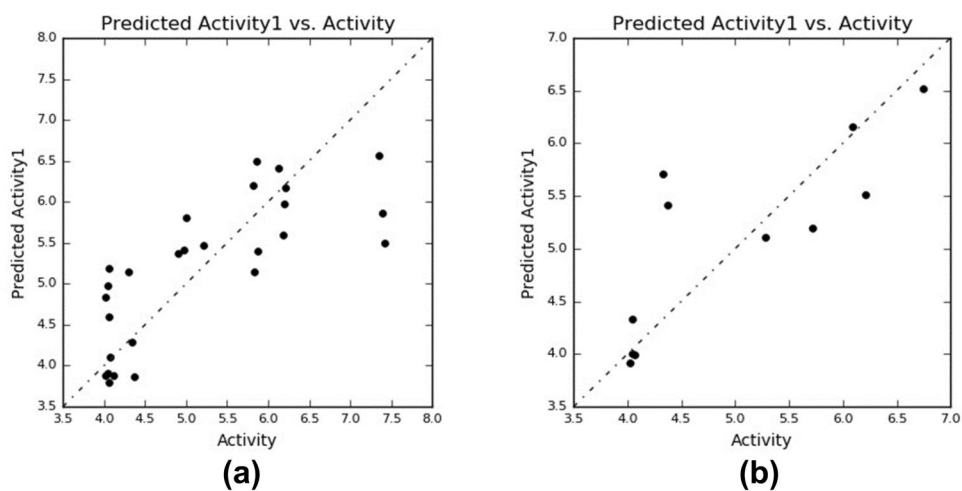


Fig. 4 A–D Visual representation of atom-based PHASE 3D-QSAR model. **A** Electron withdrawing, **B** hydrogen bond donor, **C** hydrophobic, **D** Negative, and **E** positive ionic, blue color cubes indicate

positive coefficient or increase in activity and red color cubes indicate negative coefficient or decrease in activity

Fig. 5 Graphical presentation of actual versus predicted pIC_{50} of (a) training and (b) test set molecules for obtained field 3D-QSAR model

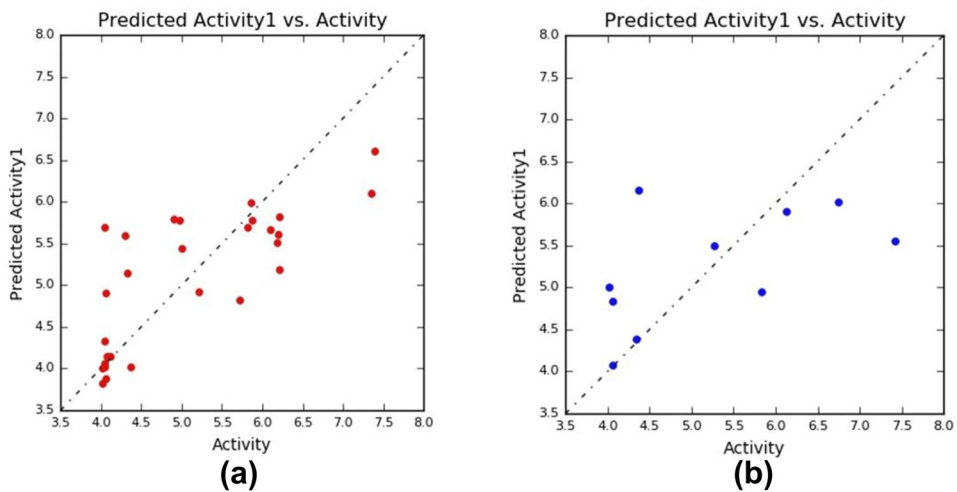
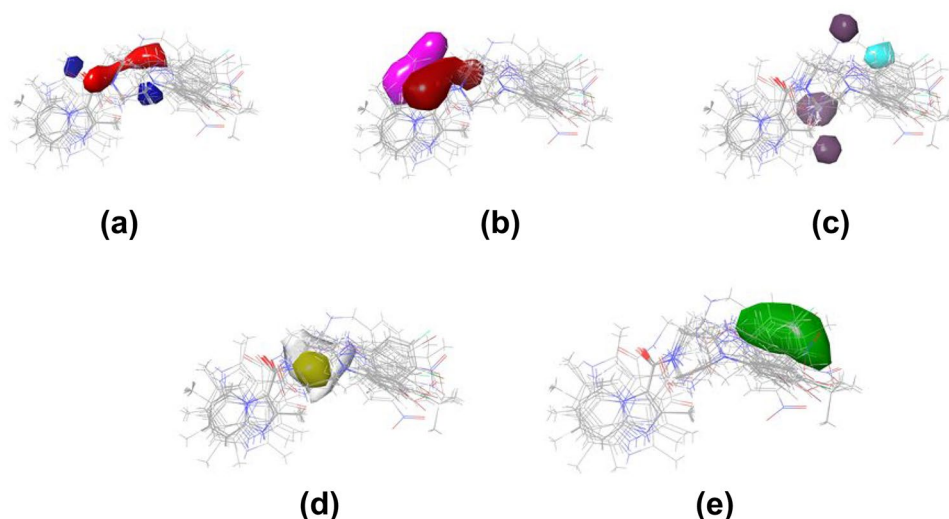


Fig. 6 Field contour maps based on test set compounds. **A** Gaussian electrostatic fields: favored electropositive (blue) and disfavored electronegative (red). **B** Gaussian hydrogen bond acceptor field: favored (red) and disfavored (magenta). **C** Gaussian hydrogen bond donor field: favored (purple) and disfavored (cyan). **D** Gaussian hydrophobic field: favoured (yellow) and disfavored (white). **E** Gaussian steric field: favored (green) unfavorable (yellow)



New York USA, 2021). For GA-MLR (genetic algorithm-based multilinear models)-based QSAR models, we used QSARINS (2.2.2 software). For in silico ADMET analysis (absorption, distribution, metabolism, excretion, and toxicity), we utilized “admetSAR” webserver [21].

Dataset, structure drawing, optimization, and molecular descriptor calculations

For this work, a dataset of thirty-eight substituted imidazopyridine (38) compounds having a wide chemical space with moderate to high anti-mycobacterium activity was selected (Table 1) [4]. All 38 imidazopyridine analogues were drawn using ChemBioDraw V.12.1. These 2D structures were then converted to 3D forms and optimized using the MMFF94 force field with the help of software, “TINKER.” For QSARINS model development, “Open3DAlign” program was used for alignments of all dataset molecules. For calculation of molecular descriptors, we used PaDEL and ChemDes. As per literature-known process, initially we divided our dataset into random splitting into 70%:30% (the training set and test set molecules, respectively). For the CPH development, we placed 27 compounds in training set and 11 molecules in the test set. Furthermore, all CPH models were assessed for their statistical significances. Throughout QSAR model developments (Tables 1–7), pIC_{50} ($pIC_{50} = -\log_{10} IC_{50}$) values are considered as dependent variables.

Pruning of molecular descriptors for GA-MLR-based QSAR models

It is very crucial to mention that descriptor pruning is the key step for the development of QSAR models. As “PaDEL” would provide more than > 30,000 molecular

descriptors, we used objective feature selection module of QSARINS ver. 2.2.2 [22–24]. Thus, many descriptors were excluded due to high co-linearity ($|r| > 0.90$) and nearly constant (> 95%) values. In continuation with the same, we also removed various esoteric descriptors manually. Finally, via descriptor pruning step, we were retained with 600 molecular descriptors (1D, 2D, and 3D).

GA-MLR-based QSAR model building

Statistically robust GA-MLR-based QSAR models were developed and validated using popular software QSARINS ver. 2.2.2. All developed QSARINS-based models were assessed for their internal and external validations, analysis of their applicability domains and Y-randomizations. These validations were carried as per the OECD guidelines. Initially, we divided all 38 molecules into training and test sets. Further, we allowed splitting of dataset as per known criteria [1, 7, 9] or method, i.e., 70%, training:30%, test, respectively. In this way, we used training set molecules for model building and test set molecules for external validations. Moreover, we also allowed multiple splitting so as to gain models with good statistical significances. During QSARINS utilization, we kept all default functions as it is. The Q^2_{LOO} was selected as a fitness function. We observed that the value of Q^2_{LOO} was increased till 6 variables and then, after there were only minor increments in Q^2_{LOO} . Thus, models with 3–6 variables were developed and assessed to check the overfitting issues. The best GA-MLR-based QSAR model was analyzed and studied further (Table 7). Further information on GA-MLR-based QSAR model validation is available in the [supporting information](#).

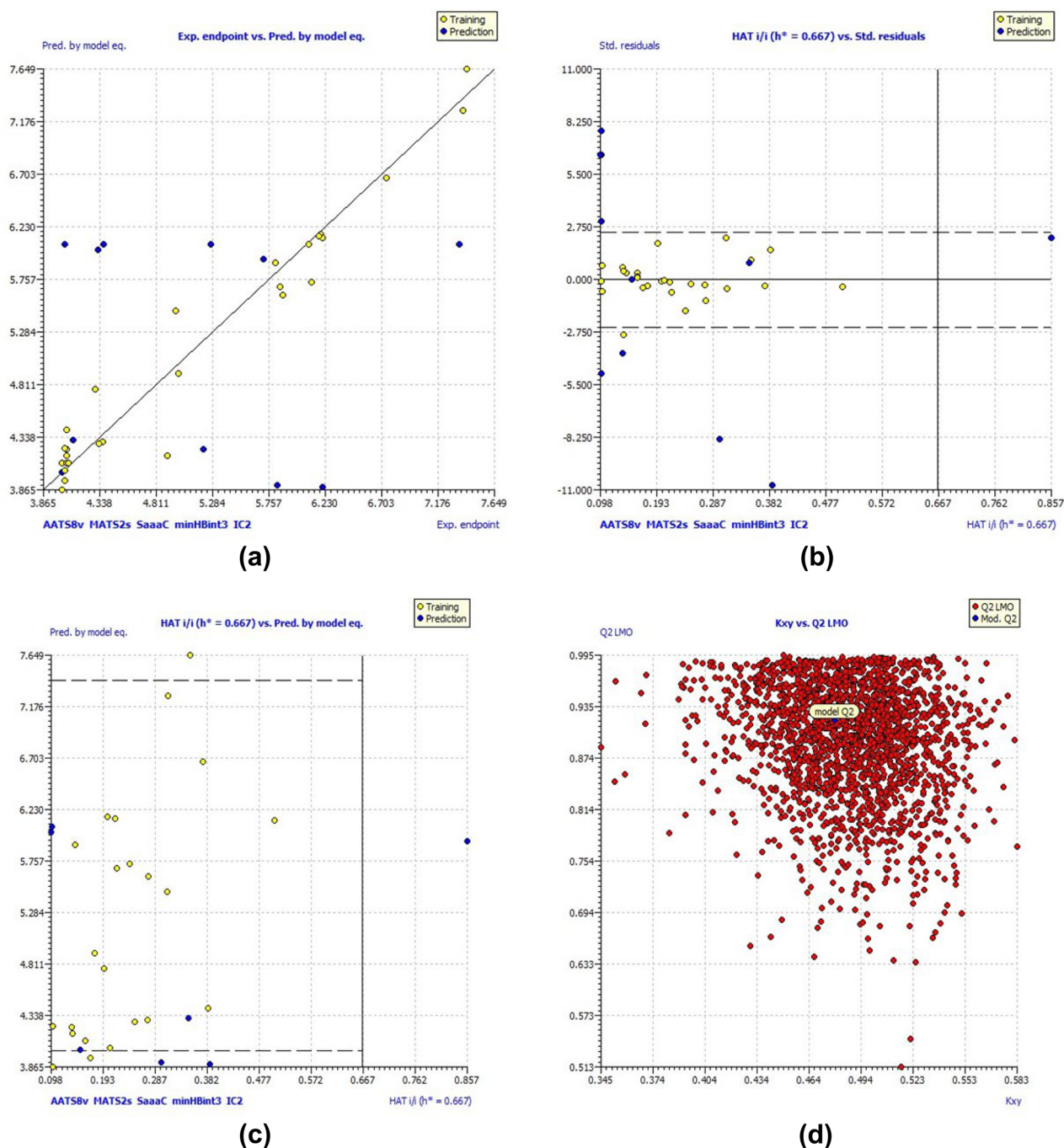


Fig. 7 **a** Graph of experimental vs predicted pIC_{50} values for model 1. **b** William's plot for model 1. **c** Insurbia plot and **d** Y-scramble plot for model 1

The CPH studies

For the developments of common pharmacophore hypotheses, we utilized two parts on our dataset molecules (38), viz., actives and inactive (Fig. 2). CPH analysis was performed with PHASE (Schrödinger, 2021 release, Inc.). The macromodel (OPLS-2005 force field) utility was used

for minimization of all 38 IMPs (imidazopyridines). Lig-Prep minimized configurations were allowed to import into PHASE workflow at $pH 7.4 \pm 0.0$. Thereafter, we set a criterion for splitting of molecules into active and inactive sets (Active = MIC (-log) > 5.40 mol/l and inactive = MIC (-log) < 4.30 mol/l). Default pharmacophoric features includes 6 features (positive (P), aromatic ring (R), negative

Table 1 Experimental dataset employed for atom-based 3D-QSAR (PLS factor 3) study along with predicted and actual pIC_{50} values (against MTB H37Rv ATCC 27,294)

Compound ID	QSAR set	Actual pIC_{50}	Predicted pIC_{50}
24	Training	7.42	6.4378
25	Test	6.208	5.63547
26a	Test	6.092	5.85956
26b	Training	5.815	5.90989
26c	Training	6.119	6.37792
26d	Training	6.194	6.20677
26e	Training	5.857	6.5087
26f	Test	6.745	6.7942
26 g	Training	7.357	6.95506
26 h	Training	7.387	7.50282
26i	Training	4.046	3.85054
26 J	Test	4.371	5.61723
26 k	Test	5.275	4.97233
26L	Test	4.325	5.69386
27A	Training	5.207	5.58198
27B	Training	5.833	5.6755
27C	Training	5.879	5.67855
27D	Training	4.06	4.24814
27E	Training	4.906	4.83818
27F	Training	4.976	5.15422
27G	Training	4.06	4.05258
28A	Training	4.304	4.37344
28B	Training	5.001	4.76742
29A	Test	5.714	5.30792
29B	Training	6.208	6.52477
30A	Test	4.02	4.25968
30B	Training	4.044	4.24333
30C	Test	4.046	4.31918
30D	Training	4.362	4.26162
30E	Training	4.114	4.29824
30F	Test	4.066	4.31047
30G	Training	4.337	4.27706
30H	Training	4.06	4.19205
30I	Training	4.074	4.12127
31	Training	6.18	5.98564
32A	Training	4.022	4.16295
32B	Training	4.022	3.65859
33	Test	4.044	4.44089

(N), hydrogen bond acceptor (A), hydrogen bond donor (D), and hydrophobic (H)). Throughout pharmacophore developments, we used default definitions of PHASE module keeping box size for pharmacophore to 2 Å. After CPH generations, top ranked CPH was selected and used further for 3D-QSAR analysis (grid space = 1 Å and PLS factor = 3).

Flexible ligand alignment and 3D-QSAR studies

For ligand alignments, we had superimposed all Lig-Prep minimized 38 ligands (LigPrep, Schrödinger, 2021). As in atom-based 3D-QSAR, a set of overlapping van der Waals spheres are considered; thus, we further carried the QSAR model development with atom-based 3D-QSAR option (Phase, Schrödinger, 2021) (Tables 1–3). Twenty-seven molecules were placed in the training set, while 11 molecules were in test set for both atom as well as field based-3D-QSAR analysis (Phase, Schrödinger, 2021) ((Tables 4–6); Figs. 3–6). The random splitting pattern (70%:30%) was used while developing atom as well as field based-3D-QSAR models. Partial least square (PLS) factor was kept as 3 along with grid spacing of 1 Å. In Field based 3D-QSAR, Gaussian based fields were employed (Fig. 6) [25]. These consists of five Gaussian-based field features like Gaussian H bond donor, Gaussian H bond acceptor, Gaussian electrostatic, Gaussian steric, and Gaussian hydrophobic. The truncate steric force field as well as electrostatic force fields were kept at 30.0 kcal/mol as per default settings. Moreover, variables with std. dev. < 0.01 were subjected to elimination. In field-based QSAR, Gaussian intensities (as descriptors) were considered as independent variables. Finally, for both atom-based and field-based 3D-QSAR models, visualization of contour maps was carried out (Figs. 3–6).

Theoretical method

The density functional theory (DFT) was utilized in order to optimize the gas-phase structures of compounds under investigations (VS-1-VS-5) [19]. The hybrid B3LYP method was used for current DFT calculations. Overall, basis set 6–311 + +G** was used while doing B3LYP calculations.

Table 2 The partial least square (PLS) statistics for atom-based 3D-QSAR models

# Factors	SD	R^2	R^2 CV	R^2 Scramble	Stability	F	P	RMSE	Q^2	Pearson R
1	0.742	0.5957	0.218	0.5467	0.856	36.8	2.42E–06	0.6	0.6348	0.8044
2	0.4571	0.8527	0.1079	0.7466	0.379	69.5	1.04E–10	0.6	0.6378	0.806
3	0.3239	0.9291	0.0927	0.8414	0.195	100.5	2.31E–13	0.63	0.5906	0.7972

R^2 value of a correlation coefficient, F variance ratio, P significance level of the variation ratio, $RMSE$ root mean square error, $Pearson R$ correlation between predicted and observed activity for test set, Q^2 value of cross-validation correlation coefficient for predicted activity

Table 3 The atom-type fraction statistics for the developed atom-based 3D-QSAR models

# Factors	H-bond donor	Hydrophobic/non-polar	Negative ionic	Positive ionic	Electron-withdrawing
1	0.069	0.617	0.028	0.065	0.221
2	0.076	0.582	0.042	0.084	0.216
3	0.081	0.583	0.048	0.088	0.201

Table 4 Experimental dataset employed for field-based 3D-QSAR (PLS factor 3) study along with predicted and actual pIC₅₀ (μM) values (against MTB H37Rv ATCC 27,294)

Ligand name	QSAR set	Activity	# Factors	Predicted activity
24	Test	7.42	1 2 3	5.99453
25	Training	6.208	1 2 3	6.31762
26a	Training	6.092	1 2 3	5.54217
26b	Training	5.815	1 2 3	5.58003
26c	Test	6.119	1 2 3	6.15684
26d	Training	6.194	1 2 3	6.21468
26e	Training	5.857	1 2 3	6.29691
26f	Test	6.745	1 2 3	6.45212
26 g	Training	7.357	1 2 3	6.703
26 h	Training	7.387	1 2 3	7.22026
26i	Training	4.046	1 2 3	4.50203
26 J	Test	4.371	1 2 3	6.18193
26 k	Test	5.275	1 2 3	5.40394
26L	Training	4.325	1 2 3	4.49476
27A	Training	5.207	1 2 3	5.57999
27B	Test	5.833	1 2 3	5.04592
27C	Training	5.879	1 2 3	5.64188
27D	Training	4.06	1 2 3	3.96404
27E	Training	4.906	1 2 3	5.25027
27F	Training	4.976	1 2 3	5.55222
27G	Test	4.06	1 2 3	4.85971
28A	Training	4.304	1 2 3	4.35371
28B	Training	5.001	1 2 3	4.49471
29A	Training	5.714	1 2 3	5.84571
29B	Training	6.208	1 2 3	6.14623
30A	Training	4.02	1 2 3	4.10262
30B	Training	4.044	1 2 3	4.11884
30C	Training	4.046	1 2 3	4.10064
30D	Training	4.362	1 2 3	4.17826
30E	Training	4.114	1 2 3	4.35093
30F	Test	4.066	1 2 3	4.12009
30G	Test	4.337	1 2 3	4.41236
30H	Training	4.06	1 2 3	3.90418
30I	Training	4.074	1 2 3	4.04521
31	Training	6.18	1 2 3	6.23863
32A	Training	4.022	1 2 3	3.53022
32B	Test	4.022	1 2 3	4.51867
33	Training	4.044	1 2 3	4.23191

The harmonic vibrational frequencies (HVF) were also calculated at same theoretical levels, after retaining of the converged geometries. The Gaussian 09 program was used for geometry optimizations and HVF analyses. Moreover, the quantum chemical descriptors derived from the conceptual DFT were calculated for the respective molecule.

In silico ADMET predictions

The “admetSAR” tool was employed to calculate ADMET properties of newer hydrazones (S1–S10). For current study, we also investigated toxicity analysis (carcinogenicities, Ames test toxicities, and CYPs enzyme–substrate/inhibitory assessments).

Results and discussion

CPH analysis, virtual screening, and 3D-QSAR

PHASE-generated pharmacophores usually consists of 6 features, which are hydrogen bond donor (D), and aromatic ring (R), hydrogen bond acceptor (A), hydrophobic group (H), negatively ionizable (N), and positively ionizable (P) (Fig. 2). Firstly, we aligned all ligands by utilizing the flexible ligand alignment with the shape-based alignment method. Ligands were classified into actives, inactive and intermediates. Finally, 5-point pharmacophore HHPRR_1 (2 hydrophobic (green), 1 positive (blue), and 2 aromatic (orange circles)) hypothesis was generated and selected based on ranking among 15 generated CPH models. CPH with HHPRR_1 feature was then visualized with excluded volumes to see features which should not clash in that region. We finally, superimposed dataset molecules (actives) with HHPRR_1 for visualization.

With the help of 3D structure of experimentally most active molecule, 24, we further used “Swiss similarity”-based ligand-based virtual screening (combined approach). From virtual screening of ZINC drug like hits, out of > 120 hits, we selected top 5 hits (VS 1–5) for further theoretical analysis and calculations of their QSAR based activity predictions. Moreover, we have also designed new series of IMPs (S1–S10) and also predicted their probable QSAR based potencies.

Table 5 The partial least square (PLS) statistics for field-based 3D-QSAR models

# Factors	SD	R^2	R^2 CV	R^2 Scramble	Stability	F	P	RMSE	Q^2	Pearson R
1	0.7054	0.5828	0.2912	0.3794	0.913	36.3	2.30E–06	0.98	0.3079	0.5551
2	0.4992	0.7991	0.1954	0.5908	0.613	49.7	1.93E–09	0.94	0.3697	0.622
3	0.3293	0.9161	0.1264	0.6834	0.339	87.3	4.73E–13	0.83	0.5039	0.7149

R^2 value of a correlation coefficient, F variance ratio, P significance level of the variation ratio, $RMSE$ root mean square error, $Pearson R$ correlation between predicted and observed activity for test set, Q^2 value of cross-validation correlation coefficient for predicted activity

Atom based- and field-based 3D-QSAR studies (statistical validations)

Table 1–6 depict statistical analysis of developed atom-based as well as field-based 3D-QSAR models. Our detailed common pharmacophoric hypothesis revealed that hydrophobic features as well as ring features are important for biological activity. Our generated plots for the training set and test set correlations with biological activity depicted acceptable 3D-QSAR equations (Fig. 3). In our current analysis of atom-based 3D-QSAR, 27 molecules were placed in training set and 11 molecules in the test set (a correlation coefficient: $R^2=0.9291$, standard deviations, $SD=0.3239$, Fischer coefficient, $F=100.5$, cross-validation correlation coefficient: $Q^2=0.5906$, $RMSE=0.63$, $Pearson R=0.7972$, $P=2.31E-13$). For the field-based 3D-QSAR, we were also retained with acceptable set of parameters as like atom-based QSAR (a correlation coefficient: $R^2=0.9161$, standard deviations, $SD=0.3293$, Fischer coefficient, $F=87.3$, cross-validation correlation coefficient: $Q^2=0.5039$, $RMSE=0.83$, $Pearson R=0.7149$, $P=4.73E-13$). Higher values for R^2 , Q^2 , and F values signifies statistical robustness of developed 3D-QSAR models. Further, these models were subjected to contour map analysis for more information of structural characteristics.

Visualizations of 3D-QSAR models

Analysis of atom-based 3D-QSAR models

From our developed PHASE-generated atom-based 3D-QSAR models, we only selected the best one with higher statistical robustness. Thus, contributions of various features were analysed by QSAR visualization. The blue occlusion maps or contours signifies increment in biological activity

(BA). However, red occlusion maps or contours suggests decrease in BA. Figure 4, represents various contour maps for different features of atom-based QSAR visualizations observed for molecule 24.

From our best generated atom-based 3D-QSAR, it was clear that electron withdrawing substitutions or groups besides a 2,6-dimethylimidazo[1,2-a]pyridine feature would increase in biological activity as represented majority by blue occlusion maps (BA). However, replacement or substitutions over amidic oxygen has been represented by red occlusion maps indicating slight decrease in BA. Hydrogen bond donating features over 2,6-dimethylimidazo[1,2-a]pyridine are well tolerated and usually would tend to increase in BA as represented with blue contours. It has also been seen that hydrophobic groups or substitutions over 2,6-dimethylimidazo[1,2-a]pyridine and 1-bromo-4-methoxybenzene feature are significantly increasing biological activities. This has been represented with large number of blue occlusion maps. Red contours for negative feature over 1-bromo-4-methoxybenzene feature indicates decrement in BA; however, positive ionic features show blue maps over the region of imidazo[1,2-a]pyridine feature. The statistical parameters and atom fractions for the developed atom-based QSAR models are tabulated in Tables 2 and 3.

Analysis of field-based 3D-QSAR models

In order to study Gaussian field-based 3D-QSAR models, we superimposed all dataset molecules over the best developed field-based 3D-QSAR model (Fig. 6). It has been observed that Gaussian electrostatic fields over amidic feature attached to 2,6-dimethylimidazo[1,2-a]pyridine prominently showed red occlusion maps indicating disfavored substitutions, while some of substitutions associated with 2,6-dimethylimidazo[1,2-a]pyridine feature indicated blue occlusions (favored pattern).

Table 6 The field type fraction statistics for the developed field-based 3D-QSAR models

# Factors	Gaussian steric	Gaussian electrostatic	Gaussian hydrophobic	Gaussian H-bond acceptor	Gaussian H-bond donor
1	0.526	0.073	0.194	0.112	0.095
2	0.459	0.093	0.201	0.13	0.117
3	0.421	0.091	0.225	0.144	0.118

Table 7 Statistical parameters for developed QSAR model 1

Statistical parameter	Model 1
Fitting	
R^2_{tr}	0.9525
R^2_{adj}	0.9411
$R^2_{tr} - R^2_{adj}$	0.0113
LOF	0.1484
Kxx	0.4239
ΔK	0.0547
RMSE _{tr}	0.2426
MAE _{tr}	0.1701
RSS _{tr}	1.5885
CCC _{tr}	0.9757
s	0.2750
F	84.2967
Internal validation	
R^2_{cv} (Q^2_{loo})	0.9190
$R^2 - R^2_{cv}$	0.0335
RMSE _{cv}	0.3168
MAE _{cv}	0.2239
PRESS _{cv}	2.7096
CCC _{cv}	0.9585
Q^2_{LMO}	0.9084
R^2_{Yscr}	0.1930
Q^2_{Yscr}	-0.3579
External validation	
RMSE _{ex}	1.4240
MAE _{ex}	1.1972
PRESS _{ext}	22.3052
R^2_{ex}	0.4000
$Q^2 - F^1$	-0.9155
$Q^2 - F^2$	-0.9158
$Q^2 - F^3$	-0.6358
CCC _{ex}	0.0046
Calc. external data regr. angle from diagonal	-44.7°
$R^2 - ExPy$ (predictions by LOO)	0.9194
R'^2_o	0.9151
k'	0.9980
r'^2_m	0.8592
R_o^2	0.9191
k	0.9983
r^2_m	0.0003

The statistical quality and strength of a GA-MLR-based QSAR model was determined on the basis of (a) internal validation based on leave-one-out (LOO) and leave-many-out (LMO) procedure (i.e., cross-validation (CV)); (b) using external validation; (c) Y-randomization (or Y-scrambling), and (d) fulfilling of respective threshold value for the statistical parameters: $R^2_{tr} \geq 0.6$, $Q^2_{loo} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $R^2 > Q^2$, $R^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, $CCC \geq 0.80$, $r^2_m \geq 0.6$, $(1 - r^2/r_o^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1 - r^2/r_o^2) < 0.1$, $0.9 \leq k' \leq 1.1$, $|r_o^2 - r_o'^2| < 0.3$ with $RMSE$ and MAE close to zero

Gaussian hydrogen bond acceptor field feature depicted prominent favored (red) and disfavored (magenta) occlusions immediate to portion attached with 2,6-dimethylimidazo[1,2-a]pyridine, which indicated intermediate effects on BA. Purple colored occlusions for the Gaussian hydrogen bond donor field indicated acceptable substitutions over a chemical bridge between 2,6-dimethylimidazo[1,2-a]pyridine moiety and 1-bromo-4-methoxybenzene ring feature. Gaussian hydrophobic field feature prominently depicted white occlusions over a chemical bridge between 2,6-dimethylimidazo[1,2-a]pyridine moiety and 1-bromo-4-methoxybenzene ring feature, thus indicating disfavored nature of substitutions with respect to BA. Green colored occlusions for the Gaussian Steric field feature over 1-bromo-4-methoxybenzene ring feature simply indicates tolerable steric features. The statistical parameters and field fractions for the developed field-based models are tabulated in Tables 5 and 6.

The analysis and interpretation of QSARINS model

For our present study, we used a small number of dataset molecules; from our previous study, it was cleared that QSAR modelling can be performed if there is a sufficient chemical space covered by analogues. It is also important to note that our currently developed model satisfies OECD (The Organisation for Economic Co-operation and Development) criteria (Table 7). Moreover, robust statistical validation parameters are also satisfying standard criteria (high values of Golbraikh and Tropsha criteria and CCCex). Thus, both internal and external validation parameters have been recorded for currently developed model and found to be statistically robust (see [supplementary information](#)) (Fig. 7). Further information on applicability domain analysis and model selection basis has been attached in [supplementary information](#).

From our analysis, we finalized our best developed models as shown below:

Multivariate model

Appendix A. Model 1 (70% training: 30% test set, 5 parametric)

Appendix B. $pIC_{50} = 10.0372 (\pm 3.8874) + 0.0517 (\pm 0.0126) * AATS8v + 23.0654 (\pm 1.8898) * MATS2s - 8.8033 (\pm 2.7514) * SaaC + 1.34 (\pm 0.1022) * minHBint3 - 1.8587 (\pm 0.4181) * IC2$

QSAR model interpretation

From our detailed analysis for QSARINS based model, it was observed that positive increments in the autocorrelation

Table 8 A dataset of newer imidazopyridine hydrazone derivatives S1–S10 (10) analogues used in the current study

Compd.	Chemical Structures	pIC_{50} (μM) values (against MTB H37Rv ATCC 27294)*	Compd.	Chemical Structures	pIC_{50} (μM) values (against MTB H37Rv ATCC 27294)
1.		12.55	6.		13.139
2.		12.53	7.		12.745
3.		12.52	8.		12.531
4.		12.53	9.		12.83
5.		11.954	10.		13.64

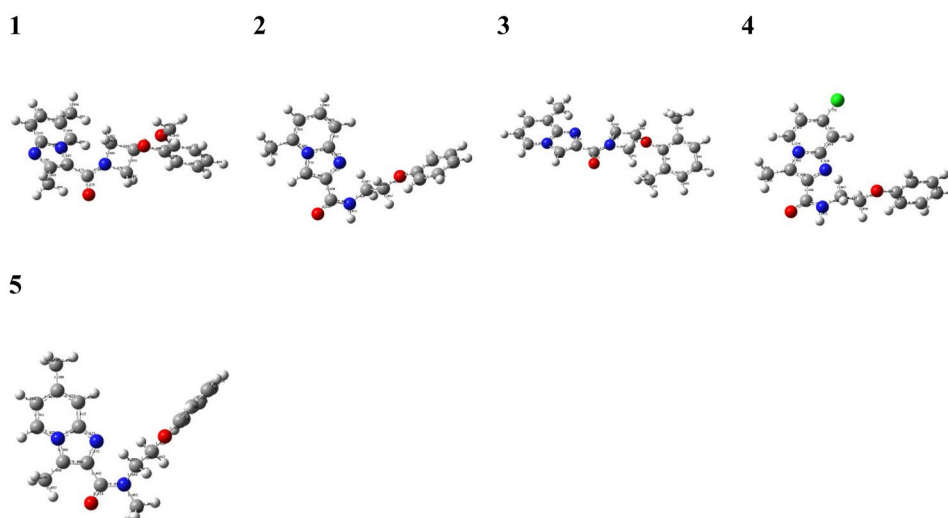
Predicted pEC_{50} values are based on model 1 QSARINS equation

Table 9 A dataset of virtually screened and theoretically studied imidazopyridine derivatives 1–5 (5) analogues used in the current study

ZINC Compd.	Chemical Structures	pIC_{50} (μM) values (against MTB H37Rv ATCC 27294)*	Compd.	Chemical Structures	pIC_{50} (μM) values (against MTB H37Rv ATCC 27294)
1.		2.670160 613	4.		7.9605
2.		7.949446 266	5.		3.128135536
3.		5.944399 335			

Predicted pEC_{50} values are based on model 1 QSARINS equation

Fig. 8 The B3LYP-optimized geometries of **1**–**5** (bond lengths in Å) (VS-1–5)



factor matters with biological activity. Similar trends have been observed for increase in descriptor values of MATS2s (Moran autocorrelation—lag 2/weighted by I-state), SaaaC (atom-type electrotopological state descriptor, i.e., sum of atom-type E-state:::C:), and minHBint3 (atom-type electrotopological state/minimum E-state descriptors of strength for potential hydrogen bonds of path length 3). However, there is decrement in biological activity as there would be decrease in descriptor values of information content index (neighborhood symmetry of 2-order) (IC2).

Considering limitations of our developed model, we believe that with the use of large pool of descriptors and large number of dataset molecules, this model can be further developed and used for designing and prediction of newer analogues of imidazopyridines as anti-TB agents. The MLR model was then applied to molecules VS 1–5 and S1-S10 (Tables 8 and 9).

The DFT studies of the virtually screened ZINC drug-like hits/compounds

Calculations of theoretical properties (the (FMO approach)

One must take into considerations several aspects of the frontier molecular orbital theory (FMO), especially HOMO and LUMO (the highest occupied and lowest unoccupied molecular orbitals) [19]. For electrophilic and nucleophilic sites determinations, one must consider LUMO and HOMO orbitals, respectively. The FMOs of the title compounds (VS-1 to VS-5) were studied in this study (Table 10). As depicted in Fig. 9, the transition from HOMO to LUMO for VS-3 and VS-5 belong to the π^* transition while that for VS-1, 2, and 4 is a charge transfer. The B3LYP-converged geometries of the studied compounds were summarized in Fig. 8.

The energy of the highest-occupied molecular orbital (E_{HOMO}), the energy of the lowest-unoccupied molecular orbital (E_{LUMO}), dipole moment (D), and the quantum chemical descriptors including the chemical potential (μ), chemical hardness (η), softness (S), and electrophilicity index (ω) calculated by the following equations:

$$\mu = \frac{I + A}{2} \quad (1)$$

$$\eta = \frac{I - A}{2} \quad (2)$$

$$S = \frac{1}{2\eta} \quad (3)$$

$$\omega = \frac{\mu^2}{2\eta} \quad (4)$$

where I and A are the ionization energy and electron affinity of a species, respectively. ²⁶

Furthermore, the ionization energy (I) and the electron affinity (A) of a species could be calculated by applying the Koopmans' theory [19] ($I = -E_{\text{HOMO}}$ and $A = -E_{\text{LUMO}}$) and the quantum chemical descriptors were calculated and summarized in the Table 10. It is pertinent to note that both chemical hardness and global softness values are comparable with the stability and reactivity of molecules. A Smaller (greater) the value of hardness (softness) a molecule has, the more reactive it should be. From Table 10, it was observed that VS-4 exhibited the lowest value of chemical hardness and highest value of global softness among the studied compounds, therefore, it is chemically more reactive and less stable than all other compounds. Moreover, from previous literature analyses, one can compare electrophilicity index (ω) with the

Fig. 9 The HOMO and LUMO of the studied compounds VS-1 to VS-5 (1–5) (the isovalue = 0.02 a.u.)

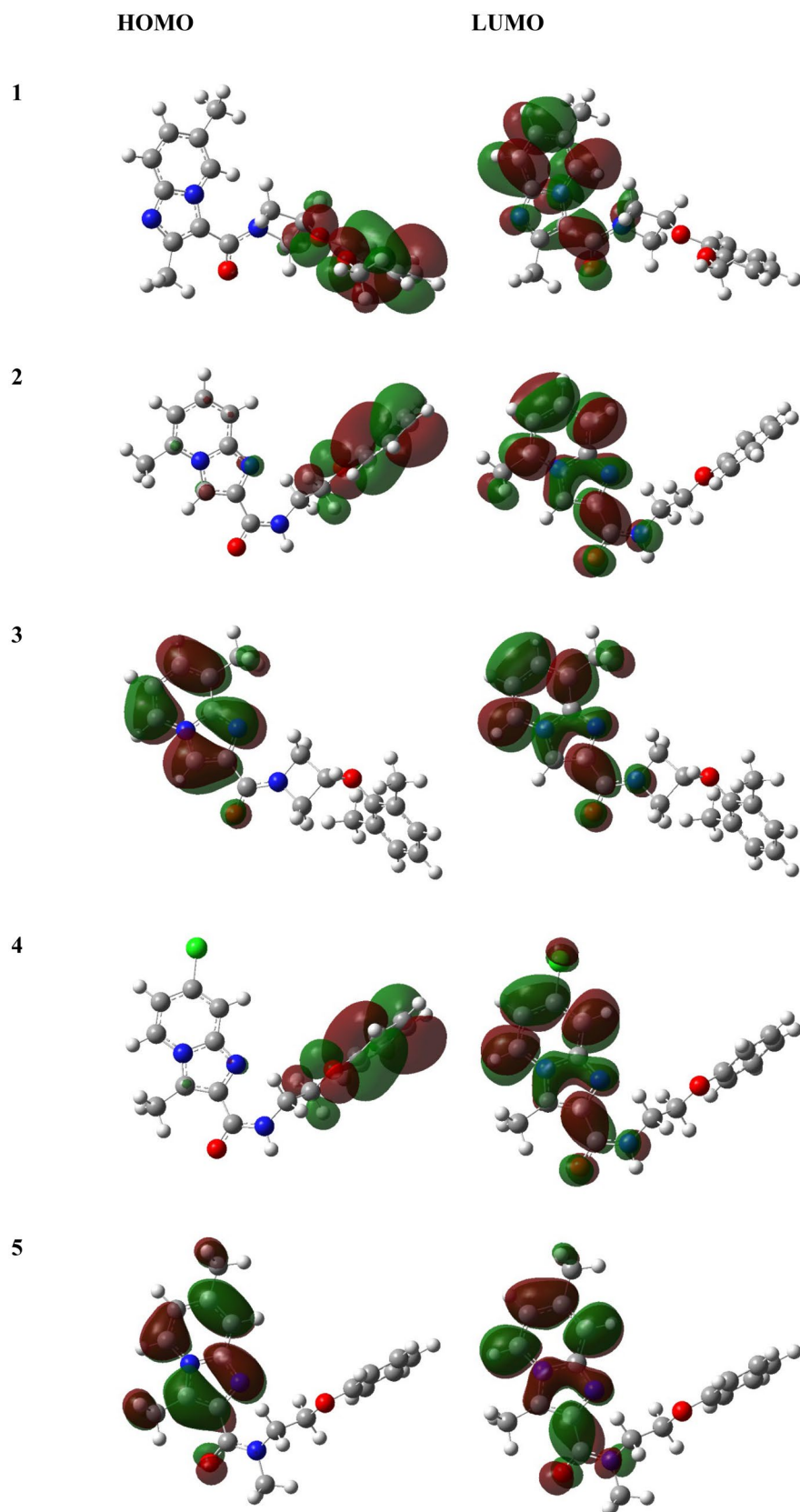


Fig. 10 The MEP of the studied compounds VS-1 to VS-5 (**1–5**) (the isovalue = 0.0004 a.u)

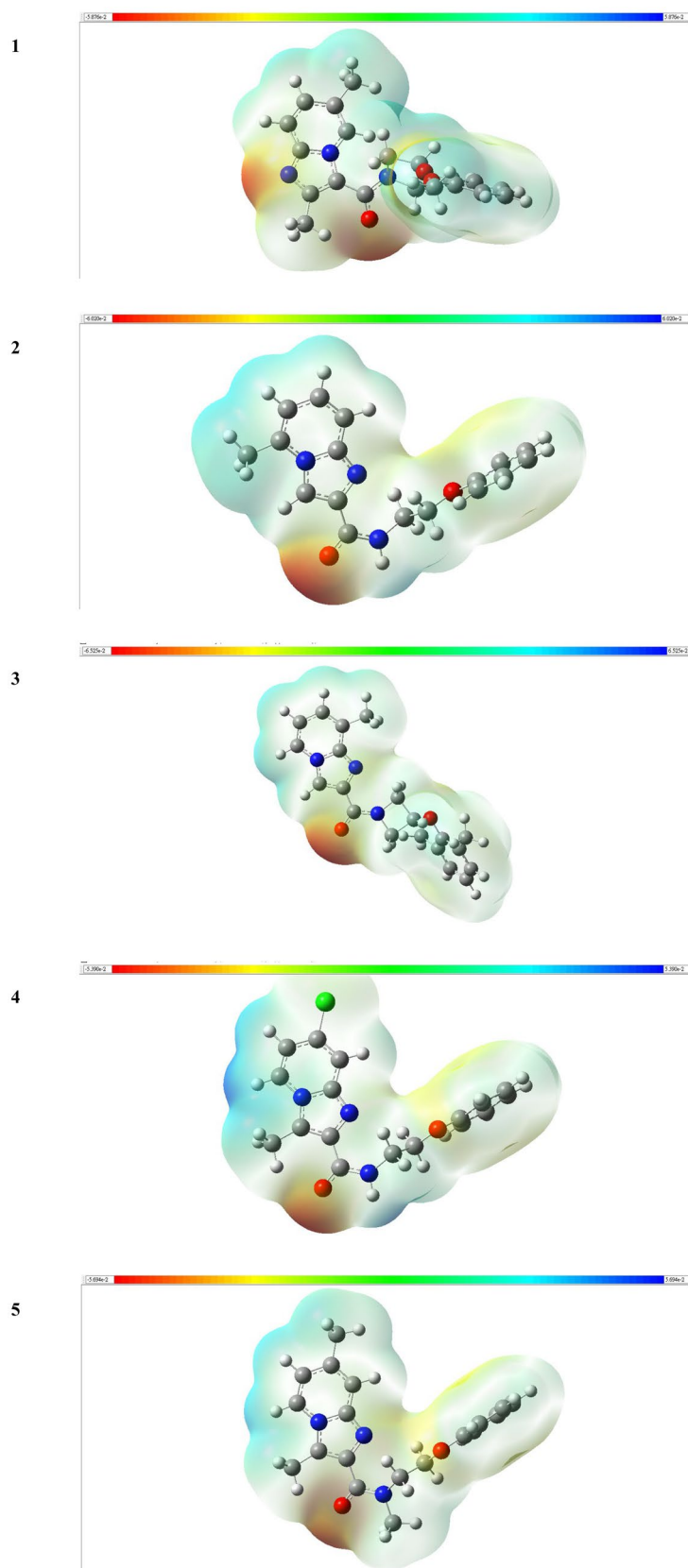


Table 10 The quantum chemical descriptors calculated for a dataset of virtually screened imidazopyridine derivatives **1–5 (5)** analogues used in the current study

ZINC comp	E_{HOMO} (eV)	E_{LUMO} (eV)	D (Debye)	μ (eV)	η (eV)	S (eV^{-1})	ω (eV)
1	−6.0185	−1.4737	5.6347	3.7461	2.2724	0.4401	3.0878
2	−6.1409	−1.6516	4.2727	3.8963	2.2447	0.4455	3.3815
3	−6.1472	−1.5085	3.7313	3.8279	2.3193	0.4312	3.1588
4	− 6.1796	− 1.9701	2.7884	4.0748	2.1047	0.4751	3.9445
5	−5.9443	−1.5099	3.9738	3.7271	2.2172	0.4510	3.1326

toxicity of molecules. From Table 10, it was observed that VS-1 showed the lowest value of electrophilicity index (ω) among the tested compounds, which indicates that it should have the lowest toxicity among all the studied compounds.

MEPs

It has been seen that molecular electrostatic potentials (MEPs) can be successfully used for measuring the interaction strengths of the nearby charges, nuclei and electrons, etc. These plots give us visual representations and provide information on the charge distributions [26]. In general, red color contours denote the lowest electrostatic potential value, while blue indicates the highest electrostatic potential value. From Fig. 10, it has been seen that oxygen atoms in studied compound are responsible for the nucleophilic attacks (as they have larger electronegativities).

Theoretical prediction of ADMET properties

Calculations of ADMET properties are very crucial to optimize lead molecules. Although, in silico methodologies have known limitations, still these methods can successfully use to predict pharmacokinetic properties before actual experiments. For our current study, we have accessed these properties via in-silico methods for virtually screened ZINC Druglike hits (VS1-5) and newly designed S1-S10 molecules (the “admetSAR” tool). Virtually screened hits were obeyed drug like characteristics. It was observed that designed molecules S1-S10 followed a class III acute oral toxicity. Moreover, eye erosion, HERG and corrosion properties were found to be negative. Our predicted hit molecule, S10 was found to be good in-silico pharmacokinetic properties, thus we would recommend molecule S10 for further in vitro analysis for future scope on this work.

Conclusion

In summary, a dataset of thirty-eight substituted imidazo[1,2-a]pyridine-3-carboxamide (38) compounds was used to develop a common pharmacophore hypothesis, atom-based as well as field-based 3D-QSAR analysis. Moreover, same dataset has been explored for GA-MLR QSAR model developments. From pharmacophoric hypothesis, HHPRR we understood that

hydrophobicity and ring functionality are key important features. For both atom-based (a correlation coefficient: $R^2=0.9291$, standard deviations, $SD=0.3239$, Fischer coefficient, $F=100.5$, cross-validation correlation coefficient: $Q^2=0.5906$, $RMSE=0.63$, Pearson $R=0.7972$, $P=2.31E-13$) and field-based (a correlation coefficient: $R^2=0.9161$, standard deviations, $SD=0.3293$, Fischer coefficient, $F=87.3$, cross-validation correlation coefficient: $Q^2=0.5039$, $RMSE=0.83$, Pearson $R=0.7149$, $P=4.73E-13$) 3D-QSAR analysis, we were retained with higher values for both internal as well as external validation parameters. Thus, by utilizing CPH and 3D-QSAR visualization along with GA-MLR-based QSAR model, one can design newer analogues with better predictivity for biological activity and thus, would act as a reliable tool. Moreover, from our ZINC drug-like screening for actual experimental hit molecule, we retained with VS 1–5 hits which were further screened for their DFT properties. We also designed a new series of IMP-hydrazone molecules (S1-S10) and studied for in-silico ADMET analysis.

From the analysis of in-silico ADMET analysis for both VS1-5 and S1-S10, molecule S10 were found to be more potent (predicted pIC_{50} (μM) value: 13.64 (against MTB H37Rv ATCC 27,294)) as well as good ADMET properties. Henceforth, the designed analogue S10 and virtually screened VS-4 molecules (predicted pIC_{50} (μM) value: 7.96 (against MTB H37Rv ATCC 27,294)) would be proposed forward as potent anti-mycobacterial agents from our combined theoretical analysis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11224-022-01879-2>.

Acknowledgements We, the author(s) clearly indicate that some of explanations for theoretical section of this manuscript (data) has been presented previously in the conference proceeding (Section 3 (subsection 3.5 (Paragraph nos. 2–4) which can be accessed at <https://sciforum.net/paper/view/11655> (13 November 2021 by MDPI in The 25th International Electronic Conference on Synthetic Organic Chemistry session Computational Chemistry).

We wish to thank the Dept. of Pharmaceutical Sciences, *Birla Institute of Technology, Mesra, India*, for financial assistance and *Schrodinger team, Bangalore, India*, for providing the technical and software support to carry out the work (a trial license provided to SM). Author is also thankful for the provision of IRF (PHD/PH/10006/20) (Ref. No. GO/Estb/Ph.D./IRF/2020–21/) provided by BIT, Mesra, India, to S.M. One of the authors SM is thankful to Dr. Paola Gramatica for providing the software QSARINS-Chem 2.2.1

Author contribution All the authors read and approved the final manuscript. SM and AP: writing, designing, visualization; BT: designing and software; CL: software and interpretations.

Funding Financial assistance provided by Dept. of Pharmaceutical Sciences, Birla Institute of Technology, Mesra, India.

Data availability statement Data will be made available upon reasonable request.

Declarations

Ethical approval None to report.

Conflict of interest The authors declare no competing interests.

References

- Mali SN, Chaudhari HK (2018) Computational studies on imidazo [1, 2-a] pyridine-3-carboxamide analogues as antimycobacterial agents: Common pharmacophore generation, atom-based 3D-QSAR, molecular dynamics simulation, QikProp, molecular docking and prime MMGBSA approaches. *Open Pharm Sci J* 5(1). <https://doi.org/10.2174/1874844901805010012>
- Tuberculosis - WHO (World Health Organization) Factsheet, Accessed online on: 21-01-2022. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis#>
- Alemu MA, Yesuf A, Girma F, Adugna F, Melak K, Biru M, Seyoum M, Abiye T (2021) Impact of HIV-AIDS on tuberculosis treatment outcome in Southern Ethiopia—a retrospective cohort study. *J Clin Tuberc Other Mycobact Dis* 25:100279
- Lv K, Li L, Wang B et al (2017) Design, synthesis and antimycobacterial activity of novel imidazo[1,2-a]pyridine-3-carboxamide derivatives. *Eur J Med Chem* 137:117–125
- Khatun S, Singh A, Bader GN, Sofi FA (2021) Imidazopyridine, a promising scaffold with potential medicinal applications and structural activity relationship (SAR): recent advances. *J Biomol Struct Dyn* 1–24
- Kshatriya R, Kambale D, Mali S, Jejurkar VP, Lokhande P, Chaudhari HK, Saha SS (2019) Brønsted acid catalyzed domino synthesis of functionalized 4H-chromens and their ADMET, molecular docking and antibacterial studies. *ChemistrySelect* 4(27):7943–7948. <https://doi.org/10.1002/slct.201901775>
- Jejurkar VP, Mali SN, Kshatriya R, Chaudhari HK, Saha S (2019) Synthesis, antimicrobial screening and in silico appraisal of iminocarbazole derivatives. *ChemistrySelect* 4(32):9470–9475. <https://doi.org/10.1002/slct.201901890>
- Jadhav BS, Yamgar RS, Kenny RS, Mali SN, Chaudhari HK, Mandewale MC (2020) Synthesis, in silico and biological studies of thiazolyl-2h-chromen-2-one derivatives as potent antitubercular agents. *Curr Comput Aided Drug Des* 16(5):511–522. <https://doi.org/10.2174/1386207322666190722162100>
- Desale VJ, Mali SN, Chaudhari HK, Thorat BR, Yamgar RS (2020) Synthesis and Anti-mycobacterium Study on Halo-substituted 2-aryl oxyacetohydrazones. *Curr Comput Aided Drug Des* 16(5):618–628. <https://doi.org/10.2174/1573409915666191018120611>
- Anuse DG, Mali SN, Thorat BR, Yamgar RS, Chaudhari HK (2020) Synthesis, SAR, In silico appraisal and anti-microbial study of substituted 2-aminobenzothiazoles derivatives. *Curr Comput Aided Drug Des* 16(6):802–813. <https://doi.org/10.2174/1573409915666191210125647>
- Thorat BR, Rani D, Yamgar RS, Mali SN (2020) Synthesis, spectroscopic, in-vitro and computational analysis of hydrazones as potential antituberculosis agents:(part-I). *Comb Chem High Throughput Screening* 23(5):392–401. <https://doi.org/10.2174/1386207323999200325125858>
- Leão RP, Cruz JV, da Costa GV, Cruz JN, Ferreira EFB, Silva RC, de Lima LR, Borges RS, dos Santos GB, Santos CBR (2020) Identification of new rofecoxib-based cyclooxygenase-2 inhibitors: a bioinformatics approach. *Pharmaceuticals* 13:209. <https://doi.org/10.3390/ph13090209>
- Santana de Oliveira M, da Cruz JN, Almeida da Costa W, Silva SG, Brito MdP, de Menezes SAF, de Jesus Chaves Neto AM, de Aguiar Andrade EH, de Carvalho Junior RN (2020) Chemical Composition, Antimicrobial Properties of Siparuna guianensis Essential Oil and a Molecular Docking and Dynamics Molecular Study of its Major Chemical Constituent. *Molecules* 25(17):3852. <https://doi.org/10.3390/molecules25173852>
- Araújo PHF, Ramos RS, da Cruz JN, Silva SG, Ferreira EFB, de Lima LR, Macêdo WJC, Espejo-Román JM, Campos JM, Santos CBR (2020) Identification of potential COX-2 inhibitors for the treatment of inflammatory diseases using molecular modeling approaches. *Molecules* 25(18):4183. <https://doi.org/10.3390/molecules25184183>
- Neto RDAM, Santos CBR, Henriques SVC, Machado LDO, Cruz JN, da Silva CHTP, Federico LB, de Oliveira EHC, de Souza MPC, da Silva PNB, Taft CA, Ferreira IM, Gomes MRF (2020) Novel chalcones derivatives with potential antineoplastic activity investigated by docking and molecular dynamics simulations. *J Biomol Struct Dyn* <https://doi.org/10.1080/07391102.2020.1839562>
- Santana de Oliveira M, Pereira da Silva VM, Cantão Freitas L, Gomes Silva S, Nevez Cruz J, de Aguiar Andrade EH (2021) Extraction yield, chemical composition, preliminary toxicity of *Bignonia nocturna* (Bignoniaceae) essential oil and in silico evaluation of the interaction. *Chem Biodivers* 18(4):e2000982
- Costa EB, Silva RC, Espejo-Román JM, de A. Neto MF, Cruz JN, Leite FHA, Silva CHTP, Pinheiro JC, Macêdo WJC, Santos CBR (2020) Chemometric methods in antimalarial drug design from 1,2,4,5-tetraoxanes analogues, SAR and QSAR in Environmental Research 31(9):677–695. <https://doi.org/10.1080/1062936X.2020.1803961>
- Mali SN, Pandey A (2021) Molecular modeling studies on 2, 4-disubstituted imidazopyridines as anti-malarials: atom-based 3D-QSAR, molecular docking, virtual screening, in-silico ADMET and theoretical analysis. *Journal of Computational Biophysics and Chemistry* 20(03):267–282. <https://doi.org/10.1142/S2737416521500125>
- Mishra VR, Ghanavatkar CW, Mali SN, Qureshi SI, Chaudhari HK, Sekar N (2019) Design, synthesis, antimicrobial activity and computational studies of novel azo linked substituted benzimidazole, benzoxazole and benzothiazole derivatives. *Comput Biol Chem* 78:330–337
- Mali SN, Pandey A (2021) Balanced QSAR and molecular modeling to identify structural requirements of imidazopyridine analogues as anti-infective agents against trypanosomiasis. *J Comput Biophys Chem* 1–32. <https://doi.org/10.1142/S2737416521410015>
- Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, Lee PW, Tang Y (2012) admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties
- Yap CW (2011) PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 32(7):1466–1474
- Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S (2013) QSARINS: A new software for the development, analysis, and validation of QSAR MLR models
- Chirico N, Sangion A, Gramatica P, Bertato L, Casartelli I, Papa E (2021) QSARINS-Chem standalone version: A new platform-independent software to profile chemicals for physico-chemical properties, fate, and toxicity. *J Comput Chem*

25. Raj KK, Manohar S, Talluri VR, Rawat DS (2015) Insights into activity enhancement of 4-aminoquinoline-based hybrids using atom-based and field-based QSAR studies. *Med Chem Res* 24(3):1136–1154
26. Zhang G, Chen Y, Liu F, Huang J, Li P, Wang B, Zhao W, Chen M, Xu S, Guan F, Tian M (2021) Comprehensive investigation of structural properties (X-ray diffraction, IR, Hirshfeld, MEP and FMOs) and in silico screening of potential biological activity of Euphorbia factor L1. *J Mol Struct* 1246:131237

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.