



Contents lists available at UGC-CARE

International Journal of Pharmaceutical Sciences and Drug Research

[ISSN: 0975-248X; CODEN (USA): IJPSPP]

Available online at www.ijpsdronline.com

Research Article

In-silico Design of Potent Anti-tubercular Agents containing Isatinylthiosemicarbazone Pharmacophore

K. G. Raut^{1*}, S. N. Kothawade², V. V. Pande², V. S. Wagh¹, S. S. Bole², R. B. Sumbe¹

¹Department of Pharmaceutical Chemistry, RSM's N. N. Sattha College of Pharmacy, Ahmednagar, Maharashtra, India.

²Department of Pharmaceutics, RSM's N. N. Sattha College of Pharmacy, Ahmednagar, Maharashtra, India.

ARTICLE INFO

Article history:

Received: 08 August, 2023

Revised: 19 October, 2023

Accepted: 24 October, 2023

Published: 30 November, 2023

Keywords:

Isatinylthiosemicarbazone, ICL inhibitor, QSAR, Docking, ADMET study.

DOI:

10.25004/IJPSDR.2023.150606

ABSTRACT

A recent study reveals that Isatinylthiosemicarbazone analogues inhibit mycobacterial growth by inhibiting the isocitrate lyase (ICL). Hence two-dimensional (2D), three-dimensional (3D), and group QSAR (GQSAR) studies were performed to reduce the amount of pharmacophore needed to make effective ICL inhibitors. New chemical entities (NCEs) were created based on the findings of all QSAR studies. It was discovered that QSAR models produced noticeably positive statistical findings. i.e. ($r^2 > 0.7$), cross-validation ($q^2 > 0.6$), and external validation ($\text{pred}_r^2 > 0.6$), indicating high predictability of all models. Utilizing molecular docking studies, the binding affinities of designed NCEs were investigated for the ICL enzyme (PDB code: 1F8M). The absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of designed NCEs were expected to have a pharmacokinetic profile similar to their drug. Overall, it is important to state that the methodology used for pharmacophore optimization using 2D, 3D, G-QSAR, and molecular docking ADMET research works was discovered to be extremely accurate.

INTRODUCTION

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* and, to a lesser extent, *M. bovis* and *M. africanum*, is still a major global health concern. The worrisome number of reported cases, estimated at 9.4 million worldwide, demonstrates its far-reaching impact. The large burden suffered in India, with an estimated 1.6–2.4 million cases, is particularly remarkable, accounting for a significant fraction of the global TB burden.^[1,2] Illness has historically been the leading cause of morbidity and mortality in infectious diseases. Research and development initiatives are required to tackle these difficulties to generate new, powerful chemical entities.

According to Pepper *et al.*'s thorough research, the treatment of patients co-infected with TB and HIV frequently entails a combination of antiretroviral and

anti-tubercular medicines. Paradoxically, as a result of the restoration of immunological function, a phenomenon known as paradoxical worsening of tubercular diseases might develop throughout the course of such treatment. Because of the relationship between drug-drug interactions and the underlying condition, this phenomenon creates unforeseen complications.^[3,4] The increased risk of tubercle bacilli remaining dormant in the lungs of people exposed to them can result in tuberculosis reactivation. The latent/dormant form of tuberculosis is resistant to both immunisation and existing anti-tubercular therapy.^[5] According to Sacchetini *et al.* in 2000, isocitrate lyase (ICL), a crucial enzyme in the glyoxalate shunt, has been related to MTB survival in macrophages and its capacity to elude the immunological response.^[6] Isatin derivatives have a long history of usage as anti-viral medicines, particularly

*Corresponding Author: Mr. K. G. Raut

Address: Department of Pharmaceutical Chemistry, RSM's N. N. Sattha College of Pharmacy, Ahmednagar, Maharashtra, India.

Email ✉: rautkunal91@gmail.com

Tel.: +91-9403742488

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2023 K. G. Raut *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution- NonCommercial-ShareAlike 4.0 International License which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

in the treatment of smallpox,^[7] vaccinia,^[8] rhinovirus,^[9] moloney leukemia,^[10] and SARS viruses.^[11]

A fundamental goal in the current drug design environment is the development of cost-effective techniques to improve efficiency and reduce labor costs and time associated with blind synthesis processes. The fundamental goal is to use existing synthesized molecules to create innovative, powerful, and selective drug-like compounds with physiologically active characteristics. The molecular modeling study method is used to compress a library of unbelievably many random compounds into a more manageable list of inhibitors with the potential to be effective. As a result, we have focused our efforts on molecular modeling of anti-tubercular medicines containing isatin thiosemicarbazone nuclei to improve efficacy against the ICL enzyme. Quantitative structure-activity relationship (QSAR) investigations were initially conducted to pursue this goal and optimize the pharmacophore requirements. QSAR investigations reveal the key structural elements required for significant biological activity, establishing links between molecules' physicochemical qualities and biological activities. The QSAR methodology is critical in predicting the biological activity of freshly developed new chemical entities (NCEs) *in-silico*.^[12] The current study examined the best methodologies for developing trustworthy and externally predictive QSAR models and useful QSAR study applications for developing NCEs with anti-tubercular action. 2D, 3D, and G-QSAR experiments were carried out using predictive QSAR modeling,^[13] the best QSAR model's output was used to generate NCEs. Using molecular modeling research, the current study successfully optimized the selected isatin thiosemicarbazone pharmacophore. Notably, the anticipated activity of the NCEs was significantly higher (KR1=11.63) than that of the most potent chemical in the previous series (B20=5.48).

MATERIALS AND METHODS

Chemical and Biological Dataset

To develop QSAR models for this investigation, 36 compounds from the isatin thiosemicarbazone series with anti-tubercular activity (Table 1) were chosen.^[14]

Molecular Modeling Tools

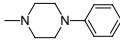
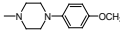
VLife Molecular Design Suite 3.5 was used for all QSAR research. The Merck molecular force field (MMFF) energy minimization method was utilized to optimize molecules.^[15]

Two-dimensional QSAR (2D QSAR) Studies

2D QSAR experimental design

A dataset of 36 molecules was manually divided into numerous training and test sets to improve the predictive capabilities and reliability of the QSAR model. Several combinations of molecules were used to establish a

Table 1: Selected series of compounds containing isatin thiosemicarbazone pharmacophore

Compound No.	R'	R ₁	Activity	
			MIC (μM)	PMIC
A7	Cl	-N(CH ₃)	76.27	4.1176
B7	Cl	-do-	18.28	4.7380
A8	F	-do-	10.05	4.9978
B8	F	-do-	9.62	5.0168
A9	CH ₃	-do-	81.34	4.0896
B9	CH ₃	-do-	19.45	4.7110
A10	Cl	-N(C ₂ H ₅) ₂	70.26	4.1532
B10	Cl	-do-	16.90	4.7721
A11	F	-do-	9.22	5.0352
B11	F	-do-	4.41	5.3555
A12	CH ₃	-do-	11.83	4.9270
B12	CH ₃	-do-	8.96	5.0476
A13	Cl		6.11	5.2139
B13	Cl	-do-	4.06	5.3914
A14	F	-do-	4.41	5.3555
B14	F	-do-	4.25	5.3716
A15	CH ₃	-do-	8.96	5.0476
B15	CH ₃	-do-	4.29	5.3675
A16	Cl		28.09	4.5514
B16	Cl	-do-	13.62	4.8658
A17	F	-do-	7.30	5.1366
B17	F	-do-	7.07	5.1505
A18	CH ₃	-do-	14.72	4.8320
B18	CH ₃	-do-	12.36	4.9079
A19	Cl	-do-	13.04	4.8847
B19	Cl	-do-	6.34	5.1979
A20	F	-do-	3.37	5.4723
B20	F	-do-	3.27	5.4854
A21	CH ₃	-do-	6.82	5.1662
B21	CH ₃	-do-	3.30	5.4814
A22	Cl		13.16	4.8807
B22	Cl	-do-	12.78	4.8934
A23	F	-do-	6.83	5.1655
B23	F	-do-	6.14	5.2118
A24	CH ₃	-do-	13.75	4.8616
B24	CH ₃	-do-	6.68	5.1752

diverse set of training and test sets, ensuring that each molecule was included in a distinct set for each cycle. A Y-randomization test was performed on the training and test sets that met all of the model evaluation criteria (Fig. 1). Y-randomization test evaluation criteria:

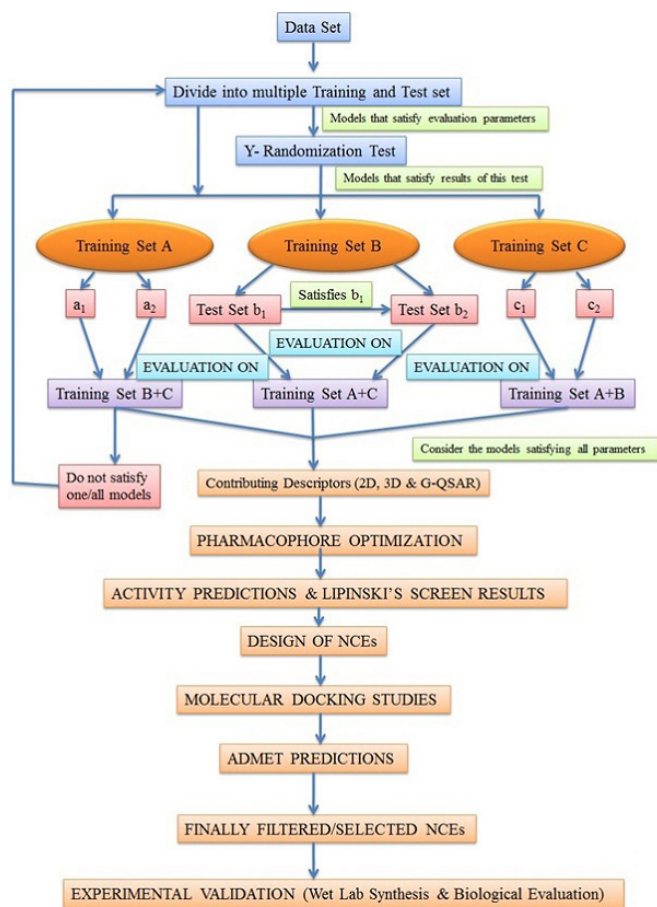


Fig. 1: Experimental design

n = number of molecules

df = Degree of freedom ($n-k-1$) (higher is better)

k = A model's number of descriptors ($\leq n/5$)

r^2 = Coefficient of determination (>0.7)

q^2 = Cross validated r^2 (>0.5)

SEE = Standard error of estimate (smaller is better)

$\text{Pred}_r^2 = r^2$ of external test set (>0.5)

F-test = In terms of the model's statistical significance, a larger value is preferable for the same descriptors and compounds.

$\text{Best_ran_}r^2$ = Y-randomization test's highest r^2 value (as low as compared to r^2)

$\text{Best_ran_}q^2$ = Y-randomization test's highest q^2 value (as low as compared to q^2)

Z score = The Y-randomization test is used to calculate it (greater is better).

Alpha = Randomization test statistical significance parameter (<0.01)

As a result of the Y-randomization test, the names training set-A, training set-B, and training set-C were given to the three selected models. To prevent chance-correlated results, these models were split into two test sets: test set a1 for the training set, test set a2 for the training set b, and test set b1 for the training set c. For the design of NCEs, only models that passed both test sets were chosen.

The following criteria were also met by the training and test sets we selected:

- A representative test set point must be close to a training set point.
- Representative points from the test and training sets must be near each other.
- The training set needs to be chemically and biologically diverse.

Uni-column statistics

Table 2 The reported findings show the statistical properties of the training and test sets prepared using the manual data selection method. The minimum and maximum values from the training and test sets must be compared so that:

- The test's maximum should be lower than the training set's maximum.
- The minimum for the test must be higher than the minimum for the practice set.

It demonstrates that the test set was interpolated or derived within the min-max range of the training set. The mean and standard deviation of the training and test sets indicate how the two sets' means and point densities (along mean) differ. The standard deviations of the training sets A, B, and C and the test sets A1, A2, and B1, B2 were found to be strikingly close. Despite their different selection in the two techniques, the data show a surprising resemblance in the distribution pattern of biological activity of molecules between the training and test sets.

Descriptor selection

528 distinct 2D descriptors were computed and pre-processed by deleting invariable columns. It has been noted in QSAR research that there is a high likelihood of chance connection between observed and predictive activity when the number of descriptors equals or surpasses the number of compounds in the dataset. As a result, decreasing the number of descriptors is an important step to avoid the inclusion of irrelevant descriptors and chance association in the final QSAR model. We used multiple descriptor selection approaches, such as forward, forward-backward, genetic algorithm, and simulated annealing, with different QSAR methodologies employing identical molecule sets to improve the model's performance and predictive capability. Following a comprehensive examination of the results, the results obtained using the forward variable selection strategy in conjunction with multiple linear regression (MLR) were finally considered.

Correlation matrix

It is a popular and necessary method for QSAR studies.^[13] The association between descriptors and activity and the correlation between descriptors and activity has been considered. We only emphasized the 2D QSAR descriptors that contributed to the selected series of chemicals and that have a direct or inverse link with activity (Table 3).



Table 2: Uni-column statistics for training sets and test sets

Parameters	Training Set B+C	Test set a_1	Test set a_2	Training Set A+C	Test set b_1	Test set b_2	Training Set A+B	Test set c_1	Test set c_2
Avg.	5.2783	5.5923	5.2889	5.0099	4.9555	5.0072	5.3630	5.1106	5.3590
Max	7.2220	6.7210	5.9830	5.4850	5.3560	5.4720	7.2220	6.0130	7.2220
Min	4.5500	4.7030	4.6130	4.0900	4.1530	4.5510	4.5890	4.5900	4.7070
S.D.	0.7059	0.7004	0.6957	0.3512	0.4212	0.3200	0.6321	0.6199	0.6119

Table 3: Correlation matrix of descriptors (2D QSAR)

S. No.	Descriptor	lpc	$T_N_N_5$	$SsFcount$	$chi2$
1	lpc	1	-	-	-
2	$T_N_N_5$	0.809744	1	-	-
3	$SsFcount$	0.000000	-	1	-
4	$chi2$	0.869211	0.971006	-	1
5	PMIC	0.7659	0.7018	0.7302	0.7296

Fitness plot

The correlation coefficient cannot disclose data that are spread between the descriptor and activity (Fig. 2). Because the correlation between descriptors and activity, rather than the distribution of data, drives variable selection methods in QSAR, certain descriptors may demonstrate chance association with activity. An accurate assessment of the fitness plot displaying the link between descriptors and activity is required to minimize such errors.

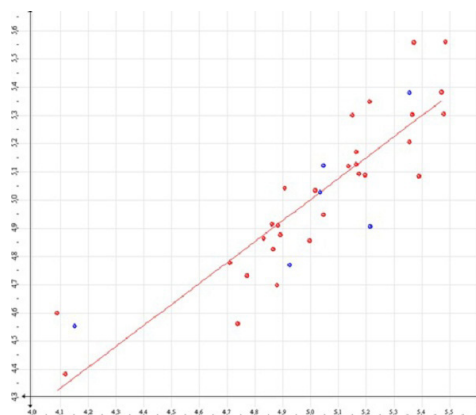
Several significant factors were considered when selecting acceptable descriptors for QSAR model construction:

The criteria for selecting descriptors for QSAR model development included achieving a balanced distribution (near to 50–50%) of data points on both sides of the best-fit line, with a slope value greater than 0.15 preferred. Furthermore, the topological descriptors were examined by examining the frequency of occurrence of various substituents within a series based on the examples found in the fitness plot. As a result, even if a descriptor has a positive association with activity and is included in the final QSAR model conclusion, its inclusion is predicated on demonstrating a well-distributed fitness plot. The careful investigation and analysis of fitness plots allowed for a reduction in the number of descriptors used.

Variance

Another key method for identifying unnecessary descriptors is to use data on descriptor variance.^[16] Our research found that several descriptors had continuous high or low variance, even with little changes in physicochemical parameters. Through rigorous investigation, we discovered that correlation, rather than variance, influences the final outcomes. As a result, more emphasis should be placed on determining the relationship between descriptions and activities.

Here is the algorithm we used to reduce the number of variables:

**Fig. 2:** Fitness plot for descriptor

- Establish the proper correlation cutoff value, denoted as A_{max} , between the descriptor and activity. Eliminate all descriptors with values lower than A_{max} .
- Establish a suitable C_{max} value for such cross-correlation cutoff between descriptors. Eliminate any descriptors with values greater than C_{max} .
- Specify the variance cutoff value for the V_{max} descriptor. Descriptors with variance values below V_{max} were eliminated.

As we discovered, this technique nearly cut the number of descriptors in half. Following that, we employed multiple regression analysis (MLR) and the manual variable selection method to ensure that each and every descriptor had a meaningful impact on the QSAR model. The goal is to reduce the number of descriptors in the final equation while still ensuring their significant contribution, which should be visible in the structural properties of the reported compounds in the sequence.

3D QSAR studies

Alignment of molecules

To confidently research 3D QSAR and practically all other areas of drug development, molecules must be accurately aligned. MolSign was used as the primary tool for post-optimization molecular alignment, allowing the discovery of both individual molecular features and shared pharmacophore features (Fig. 3).

The following color scheme is used to identify specific chemical characteristics:

Hydrogen bond donor: Magenta color

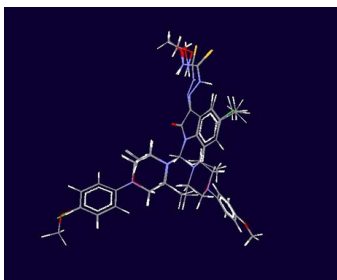


Fig. 3: Identified pharmacophore features and alignment of molecules by molsign

Hydrogen bond acceptor: Buff color

Hydrophobic: Orange colour

Aliphatic: Orange color

Negative ionizable: Green color

Positive ionizable: Violet color

The larger tessellated spheres in the representation represent the common pharmacophore features discovered in the substances, while the smaller solid components represent the individual molecules.

3D QSAR by SA-kNN-MFA

The k-Nearest Neighbour-Molecular Field Analysis (kNN-MFA) technique with the simulated annealing (SA) variable selection method was used to conduct 3D QSAR research, which has been recognized as the most relevant and applicable approach in earlier reports.^[17-19] As in the 2D QSAR study, this included creating numerous models with identical molecules in the respective training and test sets. To proceed with 3D QSAR, the specified set of molecules must be appropriately aligned following optimization, which MolSign has already performed. Atom-based alignment was used, which allowed for alignment based on the pharmacophore's individual atoms. Molecular alignment was utilised to see the composition and structure of the supplied set of molecules. A standard rectangular grid then surrounded the molecules. The energies of steric and electrostatic interactions were computed at the grid's lattice sites using a methyl probe with charge +1. These grid point interaction energy values are utilized as descriptors to calculate distances using the kNN method after being taken into account when creating relationships. The resulting associated molecules were then used to build the 3D QSAR model.

Group-based QSAR (G-QSAR) studies

Group-based QSAR (G-QSAR), a revolutionary way to investigate structure-activity connections using molecular fragments of a group of compounds, significantly improves the use of QSAR for new molecule development. By treating each molecule in the dataset as a collection of fragments, G-QSAR outperforms traditional 2D and 3D QSAR in terms of prediction capability. The fragmentation strategy, whether template-based or user-defined, allows for the evaluation of descriptors associated with each

fragment, establishing a link between these descriptors and the overall molecule's activity. Unlike traditional QSAR, G-QSAR provides us with critical site-specific recommendations regarding where to alter a given descriptor inside a molecule. G-QSAR studies employed the same training and test sets as 2D and 3D QSAR studies to develop numerous models utilising MLR analysis. The template-based fragmentation strategy and the forward variable selection method were used in the G-QSAR (Fig. 4).

Designing new chemical entities (NCEs) with the pharmacophore isatinylthiosemicarbazone.

The isatinylthiosemicarbazone pharmacophore was optimised by combining data from 2D and 3D-QSAR studies, synthesizing effective anticancer NCEs. The CombiLib tool in the VLife MDS software was used to create NCEs by taking into account the substitution pattern surrounding the pharmacophore, as shown in Fig. 6. To improve bioavailability, proposed compounds were subjected to Lipinski's screen^[20] to ensure that they had a drug-like pharmacokinetic profile. Lipinski's filters were made up of the variables listed below. (Values in parenthesis indicate ideal requirements):

- Number of hydrogen bond acceptor (A) (<10)
- Number of hydrogen bond donor (D) (<5)
- Number of rotatable bond (R) (<10)
- XlogP (X) (<5)
- Molecular weight (W) (<500 g/mol)
- Polar surface area (S) is (<140 Å)

Molecular Docking Studies

The developed compounds were exposed to molecular docking to evaluate their binding mode after displaying favorable predicted activity in all QSAR tests and complying to Lipinski's rule. For comparison, a comparative examination was performed using previously published series compounds. The compounds were then tested further to determine the best candidates based on their binding affinity compared to the typical isoniazid (INH) binding mode. GLIDE was used as the major molecular docking tool for protein-ligand docking investigations in the binding pocket of the Isocitrate lyase (ICL) enzyme in Maestro (Schrödinger Inc., USA). The crystal structures of ICL were obtained from the Protein Data Bank (ICL-PDB Code: 1F8M). The docking structures were created using the Maestro software's "Protein Preparation Wizard" and "Ligand Preparation Wizard" modules

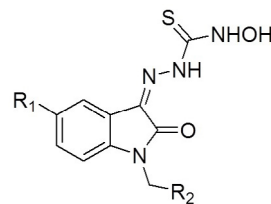


Fig. 4: Template used for G-QSAR



(Schrödinger 9.0). The refining component performed controlled affect minimization on the co-crystallized complex. This helps with the repositioning of hydroxyl groups on side chains. It employs the OPLS-AA force field for this purpose. After the co-crystallized ligand was removed from the active site, the grids were defined by centering them on the ligand in the crystal structure. The ligands were added to the project table and built with the Maestro structure builder panel. They were created with the Ligprep module, which used the MMFF94 force field to build low-energy conformers. Using the extra precision (XP) docking mode, the lowest energy conformations of the ligands were selected and docked into the grid formed from the protein structure. Unlike the ligands, the receptor remained stiff during the docking process, with the exception of a small flexibility in the protein active site. As part of the final evaluation to use the glide (docking) score, the single best pose is created as the output for each ligand.

$$G\text{-score} = a \cdot \text{vdw} + b \cdot \text{coul} + \text{Lipo} + \text{H-bond} + \text{Metal} + \text{BuryP} + \text{Rot B} + \text{Site}$$

Where,

vdW = Vander Waal energy;

Coul = Coulombic energy

Lipo = Lipophilic contact term

HBond = Hydrogen-bonding term

Metal = Metal-binding term

BuryP = Penalty for buried polar groups

RotB = Penalty for freezing rotatable bonds

Site = Polar interactions at the active site.

The coefficients of vdW and Coul are $a = 0.065$, and $b = 0.130$, respectively.

Accurately predicting protein-ligand interaction geometries is crucial for the success of structure-based drug design through virtual screening. The docking results were assessed by considering parameters such as Glide score (G-score), hydrogen bonds (H-bond), and van der Waals (vdW) interactions between the ligand and receptor.

Prediction of ADMET Properties

In some circumstances, drugs that indicate extraordinary *in-vitro* activity may later show little action or severe toxicity *in-vivo* models. Unwanted pharmacokinetic properties could be to blame for the lack of *in-vivo* activity, and reactive metabolite production could be the source of toxicity. Because NCEs failed in the later phases of the drug development process due to a lack of a drug-like pharmacokinetic profile, we had to set filters for absorption, distribution, metabolism, and excretion (ADMET) attributes. As a result, we ensured that only NCEs similar to medicines were chosen for experimental validation. All developed compounds that displayed substantial binding affinity were filtered using the Schrödinger 9.0 QikProp programme to estimate their ADMET properties. The final screening strategy was the prediction of ADMET features to find organic compounds

with both favorable anticipated activity and binding affinity with the ICL enzyme.

ADMET Prediction by QikProp, Schrödinger 9.0

Both pharmacologically and physicochemically significant descriptors were predicted. The acceptability of analogs was evaluated using Lipinski's rule of 5 to ensure a drug-like pharmacokinetic profile in rational drug design. The analogues were neutralized with QikProp before use. The OPLS-AA force field and the BOSS programme were used to create the programme, which used Monte Carlo statistical mechanics simulations on organic solutes in periodic boxes with explicit water molecules. Configurational averages were used to calculate H-bond counts and solvent-accessible surface area (SASA) descriptors. The QikProp tool used methods to simulate entire Monte Carlo simulations, yielding findings that were comparable to experimentally determined properties.

RESULTS

2D QSAR Models

One set of four meaningful descriptors was used in MLR, and chi2 results indicated that they contributed up to 48.41% to anti-tubercular activity (Fig. 5).

$$\text{MIC} = -0.000 \text{ Ipc} - 2.1052 \text{ T_N_N_5} + 0.3412 \text{ SsFcount} + 0.7057 \text{ chi2} + 0.0009$$

$r^2 = 0.7838$, $q^2 = 0.7097$, F-test = 22.65, Pred- $r^2 = 0.6172$. Two models were created from the training set A, B, and C, for each descriptor sets (Table 4).

Using 3D QSAR, the requirements for electrostatic, steric, and hydrophobic properties around the isatinylthiosemicarbazone pharmacophore were optimised. Effective NCEs were designed using property values derived from the generated data points. The ranges of data point values were determined based on the variation of field values at selected points, using the most active molecule and its nearest neighbour set as a reference. The SA-kNN-MFA 3D QSAR model produced the points E-1378 (-0.431115 -0.418058), and S-1408, (-0.173512 -0.17349) i.e., electronic and steric data points at lattice points of 1378 and 1408 respectively.

The resultant G-QSAR descriptors and model and their statistical parameters are presented in Table 5.

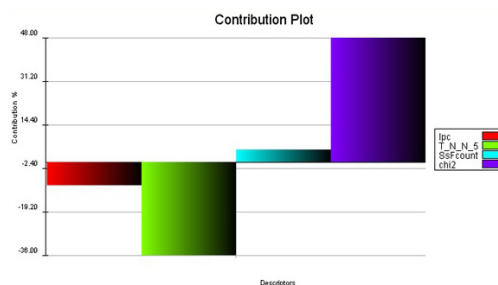


Fig. 5: Contribution plot of selected descriptors

Table 4: Statistical parameters of developed QSAR models for descriptor by forward variable selection method and MLR

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a_1	Test Set a_2	Test Set b_1	Test Set b_2	Test Set c_1	Test Set c_2
N	30	30	30	30	30	30
r^2	0.8013	0.7631	0.7838	0.8032	0.8104	0.7945
r^2 se	0.2100	0.1856	0.1759	0.2010	0.1002	0.1145
q^2	0.7075	0.6252	0.7097	0.6996	0.7271	0.7507
q^2 se	0.1012	0.1235	0.2038	0.1869	0.2045	0.1456
F-Test	23.19	18.52	22.65	23.46	24.57	20.63
Pred_ r^2	0.8697	0.9010	0.6172	0.7865	0.7234	0.8422
Pred_ r^2 se	0.2890	0.1375	0.2038	0.2619	0.2228	0.1987
	(+) VELY CONTRIBUTING			(-) VELY CONTRIBUTING		
DESCRIPTORS (Test set b_1)	SsFcount (5.74%) chi2 (48.41%)			Ipc (-9.06%) T_N_N_5 (-36.79%)		

MIC= -0.000 Ipc - 2.1052 T_N_N_5 + 0.3412 SsFcount + 0.7057 chi2 + 0.0009

Table 5: Resultant G-QSAR descriptors & model and their statistical parameters

Statistical Parameters	MLR	Contributing descriptors
N	29	Positive Negative R2-SssCH2count, R2-AlphaR R2-SsssCHcount, R1-FluorinesCount
df (n-k-1)	24	Sum(R2-SssCH2count, R2-AlphaR) Sum(R2-SsssCHcount, R1-FluorinesCount) Mult(R2-2PathCount, R1-polarizabilityAHC)
r^2	0.7466	
r^2 se	0.3671	
q^2	0.6662	
q^2 se	0.4213	
F- est	29.4617	
Pred_ r^2	0.8149	
Pred_ r^2 se	0.2963	
Best_Rand_ r^2	0.29042	
Best_Rand_ q^2	0.11536	
Z score_ r^2	11.12955	
Z score_ q^2	10.11185	
α _Rand_ r^2	0.00000	
α _Rand_ q^2	0.00000	

G-score

The G-score indicates the binding affinity of the designed compound towards the receptor or enzyme. Isoniazid (INH), a standard compound, exhibited a significant binding affinity which was found to have a G-score of -7.034, while only 4 of the 27 designed NCEs had a G-score that was higher than the standard. The designed NCEs KR1, KR2, KR3, and KR6 were found to have respective G-scores of -7.7556, -7.86310, -9.03129, and -7.7420. The binding affinity of that compound is increased when the G-score value is more negative. The detailed examination

of these findings leads to the conclusion that the designed NCEs have higher enzyme binding affinities than usual.

DISCUSSION

The results reveal that chi2 meets all of the evaluation criteria on its own. The correlation matrix confirms that it has a well-distributed pattern of data points and the strongest link with activity. Various descriptor combinations were developed to improve prediction ability while keeping chi2 and T_N_N_5 (the second highest and most adversely impacting descriptors) fixed.



Table 6: Test set B₁ & B₂ and training set A + C along with biological activity, predicted activity, and residual data

S. No.	Comp.	Biological activity (pMIC)	Predicted Activity	Residuals
		Training Set A+C		
1	A16	4.551	4.6472	-0.0962
2	A17	5.137	5.1200	0.017
3	A18	4.832	4.8635	-0.0315
4	A19	4.885	4.9092	-0.0242
5	A20	5.472	5.3823	0.0897
6	A21	5.166	5.1258	0.0402
7	A22	4.881	4.6970	0.184
8	A23	5.165	5.1699	-0.0049
9	A24	4.862	4.9136	-0.0516
10	A7	4.118	4.3825	-0.2645
11	A8	4.998	4.8553	0.1427
12	A9	4.09	4.5990	-0.509
13	B10	4.772	4.7313	0.0407
14	B11	5.356	5.2041	0.1519
15	B12	5.048	4.9476	0.1004
16	B13	5.391	5.0842	0.3068
17	B14	5.372	5.5570	-0.185
18	B15	5.367	5.3007	0.0663
19	B16	4.866	4.8257	0.0403
20	B17	5.151	5.2986	-0.1476
21	B18	4.908	5.0421	-0.1341
22	B19	5.198	5.0874	0.1106
23	B20	5.485	5.5602	-0.0752
24	B21	5.481	5.3039	-0.1771
25	B22	4.893	4.8756	0.0174
26	B23	5.212	5.3484	-0.1364
27	B24	5.175	5.0922	0.0828
28	B7	4.738	4.5610	0.177
29	B8	5.017	5.0338	-0.0168
30	B9	4.711	4.7776	-0.0666
Test set b1				
1	A10	4.153	4.5532	-0.4002
2	A11	5.035	5.0260	0.009
3	A12	4.927	4.7697	0.1573
4	A13	5.214	4.9056	0.3084
5	A14	5.356	5.3784	-0.0224
6	A15	5.048	5.1221	-0.0741

Similar calculations were carried out for Test sets A₁ & A₂ and C₁ & C₂ of set-I as well as for Set-II. In both the sets, models B₁ & B₂ (evaluated on training set A+C) were found to be the best models.

Accuracy of Model

The residual value is an important aspect in verifying the model's correctness. When the residual number is near to zero, it shows a better level of accuracy because the discrepancy between actual and projected activity is minor (≈ 0) (Table 6).

$$\text{Residuals} = \text{Actual Biological Activity (MIC)} - \text{Predicted Activity}$$

Interpretation of 2D QSAR

According to the current QSAR models, both chi2 and Baumann's alignment independent (AI) descriptors contribute significantly to explaining activity variance. T_XY_Z denotes the total number of fragments created by atom types X and Y separated by a topological Z-bond distance. The descriptors that had a substantial influence on the QSAR models are interpreted below, with percentile contribution values supplied in brackets:

- **lpc [-9.06%]:** This is a specific kind of descriptor based on information theory.
- **T_NN₅ [-36.79%]:** This represents the count of nitrogen atoms (bonded as single, double, or triple) that are separated from any other nitrogen atom (bonded as single, double, or triple) by 5 bonds within a molecule.
- **SsF count [5.74%]:** This descriptor quantifies the number of fluorine atoms connected to other atoms within a molecule through a single bond.
- **chi2 [48.41%]:** This descriptor represents a second-order retention index obtained from the gradient retention times.

After thoroughly examining the model descriptors, it was discovered that chi2, an indicator variable, has a considerable positive impact on the QSAR equation (48–49%). This emphasizes the relevance of the molecule's retention index in generating strong anti-tubercular efficacy. In contrast, the presence of -N on the ring has a negative impact on biological activity, as evidenced by descriptors such as T_NN₅, which has an inverse connection with activity. Furthermore, other features that favourably influence the design of possible anti-tubercular compounds shed light on the significance of specific atoms or groups at various places on the ring, albeit to varying degrees.

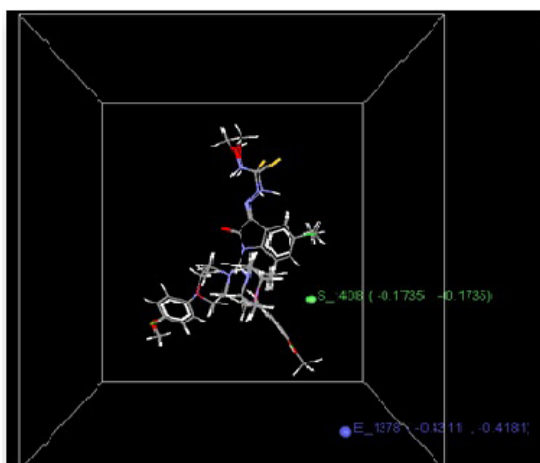
3D QSAR models

The SA-kNN-MFA method was used to pick two descriptors that matched all of the statistical parameters in the created models after rigorous investigation. Models b1 and b2 emerged as the best models based on the estimated residuals for each model. Table 7 compares the statistical data derived from the 3D QSAR generated by the SA-kNN-MFA approach.

Grid points in a 3D rectangular grid were generated using the SA-kNN-MFA approach. These grid points were then used to understand the 3D QSAR study, as shown in Fig. 6.

Table 7: Comparison of the various statistical results of 3D QSAR generated by SA-KNN-MFA method

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a1	Test Set a2	Test Set b1	Test Set b2	Test Set c1	Test Set c2
N	30	30	30	30	30	30
k-NN	2	2	2	2	2	2
q ²	0.5002	0.5231	0.5018	0.5796	0.5519	0.5180
q ² se	0.2626	0.2612	0.2554	0.2162	0.2205	0.2135
Pred_r ²	0.5321	0.5695	0.5227	0.5367	0.5822	0.5116
Pred_r ² se	0.1509	0.2026	0.2584	0.1208	0.3018	0.2762
	(+) VELY CONTRIBUTING		(-) VELY CONTRIBUTING			
DESCRIPTORS (Test Set b ₁)	-		E_1378 (-0.431115 -0.418058) S_1408 (-0.173512 -0.17349)			

**Fig. 6:** Grid points generated by SA-KNN-MFA method in 3D rectangular grid interpretation of 3D QSAR

- Electrostatic data points with negative values demonstrated the need for electronegative substituents (such as -Cl, -Br, or -OH) to increase biological activity.
- A low negative steric value range suggested that moderately bulky groups (such as C₆H₅ and CH₂C₆H₅) are needed to boost activity.

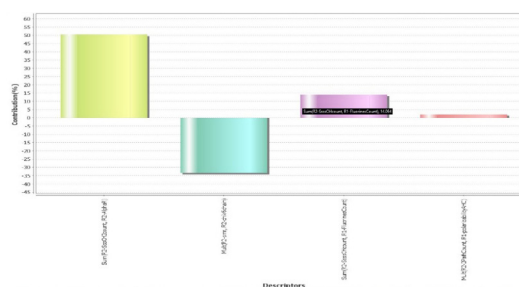
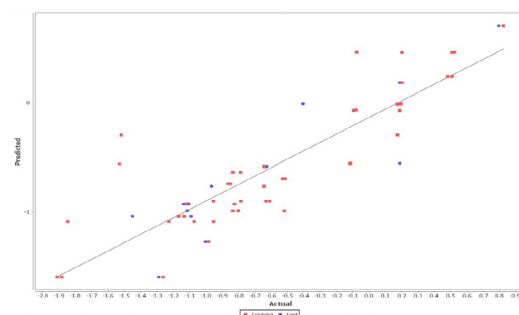
G-QSAR model

After integrating multiple QSAR approaches, six descriptors were chosen and utilized to generate MLR models that passed all statistical criteria. The model's residuals were computed, proving it the best model for 2D and 3D QSAR investigations. Fig. 7 depicts the contribution plot from the G-QSAR model.

The fitness plot for the G-QSAR model is shown in Fig. 8.

Interpretation of G-QSAR

A set of six descriptors was identified, indicating the specific positional requirements of a particular category of substituents on the pharmacophore.

**Fig. 7:** Contribution plot for G-QSAR model**Fig. 8:** Fitness plot for g-qsar model

R2-SsssCHcount- This description lists all -CH groups that are joined together by three single bonds.

R1-FluorinesCount- This descriptor represents the count of fluorine atoms present in a compound.

R2-SssCH₂count- This descriptor quantifies the total count of -CH₂ groups connected by two single bonds in a compound.

Design of NCEs with the pharmacophore *N*-phenyl-2, 2-dichloroacetamide

Based on the interpretations of 2D and 3D-QSAR data, a total of 70 newly created chemical entities (NCEs) were developed. Only 27 of these NCEs outperformed the most potent molecule in the original series in terms of projected activity, meeting the criteria of both QSAR predictions and Lipinski's screening (score 6). These 27 NCEs were chosen for future *in-silico* research. The study also discusses the



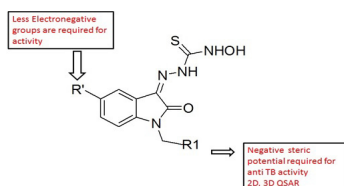


Fig. 9: Pharmacophoric requirements around isatinyl thiosemicarbazone nucleus

pharmacophoric criteria in the vicinity of the Isatinyl thiosemicarbazone nucleus in Fig. 9.

Table 8 This section presents the structural details of the designed NCEs, along with their predicted activities obtained from both 2D (using two sets of descriptors) and 3D-QSAR equations.

Molecular docking studies

It has been discovered that isatinylthiosemicarbazone analogues bind to the binding region of ICL in molecular docking results (Maestro, Schrödinger 9.0).

ICL enzyme evaluation of molecular docking results (PDB code: 1F8M)

H-bond interactions

The H-bond, a critical parameter in evaluating the activity of drug compounds, was assessed in the docking results. A comparison was made between the designed NCEs and the standard H-bond interactions in terms of quantity and length. INH, the reference compound, formed four significant H-bonds with the key binding amino acids Ser and Asn-315. Among the 27 NCEs, only four compounds, including the most potent compound B20, exhibited better H-bond results. H-bond length is another important factor in molecular docking, as shorter lengths indicate higher affinity to the target amino acid. In our study, all four compounds displayed shorter H-bonds than the reference structure. Notably, compound KR2 demonstrated the highest binding affinity with the essential amino acids in the ICL binding pocket.

Contacts

Vander Waals (vdW) interactions are used to represent the contacts.

- Good vdW interactions
- Bad vdW interactions
- Ugly vdW interactions

The analysis revealed that the 4 NCEs exhibited a higher number of favorable van der Waals (vdW) interactions, a lower number of unfavorable vdW interactions, and no unfavorable contacts when compared to INH. The results of molecular docking studies conducted with the extra precision mode of Glide are presented in the following section (Table 9).

The conformation of compound KR1 in the binding site of the ICL receptor (PDB Code: 1F8M) is shown in Fig. 10. The conformation of compound KR2 in the binding site of the ICL receptor (PDB Code: 1F8M) is shown in Fig. 11

The conformation of compound KR3 in the binding site of the ICL receptor (PDB Code: 1F8M) is shown in Fig. 12. The conformation of compound KR6 in the binding site of the ICL receptor (PDB Code: 1F8M) is shown in Fig. 13.

ADMET predictions

Lipinski's rule was applied to assess the drug-like properties of the NCEs generated by CombiLib. ADMET characteristics were also predicted, and the ranges considered suitable for a potential drug were compared. Six of the designed NCEs exhibited favorable results within the desired ranges.

ADMET prediction by QikProp, Schrödinger

Schrödinger's QikProp version 9.0 was used as the last screening technique to predict various attributes of the created analogs. Descriptors that contributed considerably to the prediction of drug-like properties were considered. The attributes are shown below, along with their matching ideal values.

- Compounds adhering to Lipinski's Rule were anticipated to possess a pharmacokinetic profile resembling that of a drug.

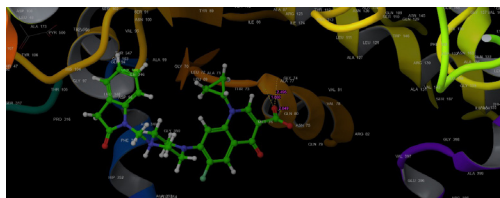


Fig. 10: The conformation of compound KR1 in the binding site of the ICL receptor (PDB code: 1F8M)

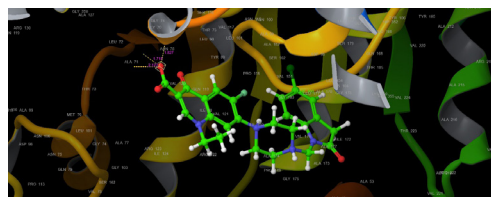


Fig. 11: The conformation of compound KR2 in the binding site of the ICL receptor (pdb code: 1F8M)

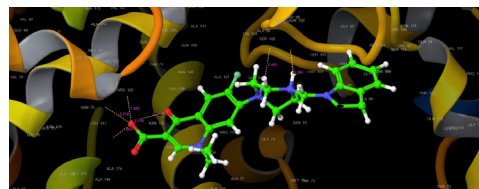


Fig. 12: The conformation of compound KR3 in the binding site of the ICL receptor (PDB code: 1F8M)

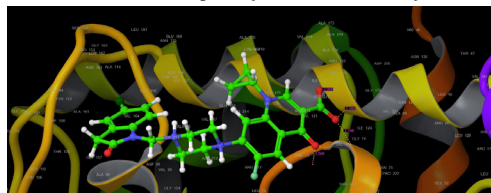


Fig. 13: The conformation of compound KR6 in the binding site of the ICL receptor (PDB code: 1F8M)