Saudi Pharmaceutical Journal 30 (2022) 693–710

Contents lists available at ScienceDirect

Saudi Pharmaceutical Journal

journal homepage: www.sciencedirect.com

Original article

Identification of potent aldose reductase inhibitors as antidiabetic (Anti-hyperglycemic) agents using QSAR based virtual Screening, molecular Docking, MD simulation and MMGBSA approaches

Ravindra L. Bakal [a], Rahul D. Jawarkar [a,*], J.V. Manwar [b], Minal S. Jaiswal [b], Arabinda Ghosh [c], Ajaykumar Gandhi [d], Magdi E.A. Zaki [e,*], Sami Al-Hussain [e,*], Abdul Samad [f], Vijay H. Masand [g], Nobendu Mukerjee [h,i], Syed Nasir Abbas Bukhari [j], Praveen Sharma [k], Israa Lewaa [l]

[a] Department of Medicinal Chemistry, Dr. Rajendra Gode Institute of Pharmacy, University-Mardi Road, Amravati, Maharashtra, India
[b] Department of Medicinal Chemistry and Pharmacognosy, Dr. Rajendra Gode College of Pharmacy, University-Mardi Road, Amravati, Maharashtra, India
[c] Microbiology Division, Department of Botany, Gauhati University, Guwahati, Assam 781014, India
[d] Department of Chemistry, Government College of Arts and Science, Aurangabad, Maharashtra 431 004, India
[e] Department of Chemistry, Faculty of Science, Al-Imam Mohammad Ibn Saud Islamic University, Riyadh 13318, Saudi Arabia
[f] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Tishk International University, Erbil, Kurdistan Region, Iraq
[g] Department of Chemistry, Vidyabharti Mahavidyalaya, Camp Road, Amravati, Maharashtra, India
[h] Department of Microbiology, Ramakrishna Mission Vivekananda Centenary College, West Bengal 700118, Kolkata, India
[i] Department of Health Sciences, Novel Global Community Educational Foundation, Australia
[j] Department of Pharmaceutical Chemistry, College of Pharmacy, Jouf University, Aljouf, Sakaka, 2014, Saudi Arabia
[k] Department of Pharmaceutics, Vinayaka College of Pharmacy, Hathod, Indore, Madhyapradesh, India
[l] Department of Business Administration, Faculty of Business Administration, Economics and Political Science, British University in Egypt, Elsherouk City, Cairo 11837, Egypt

ARTICLE INFO

Article history:
Received 23 October 2021
Accepted 1 April 2022
Available online 7 April 2022

Keywords:
QSAR
Antidiabetic
Aldose reductase
GA-MLR
Molecular Docking

ABSTRACT

The aldose reductase (AR) enzyme is an important target enzyme in the development of therapeutics against hyperglycaemia induced health complications such as retinopathy, etc. In the present study, a quantitative structure activity relationship (QSAR) evaluation of a dataset of 226 reported AR inhibitor (ARi) molecules is performed using a genetic algorithm – multi linear regression (GA-MLR) technique. Multi-criteria decision making (MCDM) analysis furnished two five variables based QSAR models with acceptably high performance reflected in various statistical parameters such as, R^2 = 0.79–0.80, Q^2_{LOO} = 0.78–0.79, Q^2_{LMO} = 0.78–0.79. The QSAR model analysis revealed some of the molecular features that play crucial role in deciding inhibitory potency of the molecule against AR such as; hydrophobic Nitrogen within 2 Å of the center of mass of the molecule, non-ring Carbon separated by three and four bonds from hydrogen bond donor atoms, number of sp2 hybridized Oxygen separated by four bonds from sp2

Abbreviations: **Ari**, Aldose Reductase Inhibitors; **QSAR**, **Quantitative** structure activity Relationship; **MD**, Molecular Dynamic; **MMGBSA**, Molecular mechanics generalized born surface area; **CADD**, **C**omputer **A**ided **D**rug **D**esigning; **SMILES**, **S**implified **M**olecular, **I**nput **L**ine-**E**ntry **S**ystem; **GA**, **G**enetic **A**lgorithm; **MLR**, **M**ultiple **L**inear **R**egression; **QSAR**, **Q**uantitative **S**tructure-**A**ctivity **R**elationship; **OLS**, **O**rdinary **L**east **S**quare; **QSARINS**, QSAR Insubria; **OECD**, **O**rganization for **E**conomic **C**o-operation and **D**evelopment; **CCC**, **C**oncordance **C**orrelation **C**oefficient; **allminus_SASA**, Solvent Accessible surface area of the all negatively charged atoms; **Minus_don_3B**, Occurrence of a donor within three bonds from a negatively charged atom; **don_ringC_6Ac**, Occurrence of the partially charged ring carbon atoms within 6Å of the donor; **don_notringC_4B**, Occurrence of the non-ring carbon atom within four bonds from the donor; **fsp2Osp2C4B**, the frequency of occurrence of a sp^2 hybridized carbon atom exactly four bonds from a sp2 hybridized oxygen atom; **com_ringCminus_2A**, Encodes information on the number of negatively charged ring Carbon atoms within 2Å from the center of mass of the molecule; **com_Nhyd_2A**, presence of hydrophobic nitrogen within 2Å of the center of mass; **H_ringN_2B**, H_ringN_2B represents a combination of ring nitrogen and hydrogen separated by within 2 bonds.

* Corresponding authors.

E-mail addresses: rlbakal@gmail.com (R.L. Bakal), rahuljawarkar@gmail.com (R.D. Jawarkar), jvmanwar@gmail.com (J.V. Manwar), jaiswalminal77@gmail.com (M.S. Jaiswal), dra.ghosh@gauhati.ac.in (A. Ghosh), gascajay18@gmail.com (A. Gandhi), Mezaki@imamu.edu.sa (M.E.A. Zaki), sahussain@imamu.edu.sa (S. Al-Hussain), abdul.samad@tiu.edu.iq (A. Samad), vijaymasand@gmail.com (V.H. Masand), nabendu21@rkmvccrahara.org (N. Mukerjee), sbukhari@ju.edu.sa (S. Nasir Abbas Bukhari), praveen140581@gmail.com (P. Sharma), israa.lewaaa@gmail.com (I. Lewaa).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jsps.2022.04.003

MDS
Virtual screening

hybridized Carbon atoms, etc. 14 in silico generated hits, using a compound 18 (a most potent ARi from present dataset with pIC_{50} = 8.04 M) as a template, on QSAR based virtual screening (QSAR-VS) furnished a scaffold 5 with better ARi activity (pIC_{50} = 8.05 M) than template compound 18. Furthermore, molecular docking of compound 18 (Docking Score = –7.91 kcal/mol) and scaffold 5 (Docking Score = –8.08 kcal/mol) against AR, divulged that they both occupy the specific pocket(s) in AR receptor binding sites through hydrogen bonding and hydrophobic interactions. Molecular dynamic simulation (MDS) and MMGBSA studies right back the docking results by revealing the fact that binding site residues interact with scaffold 5 and compound 18 to produce a stable complex similar to co-crystallized ligand's conformation. The QSAR analysis, molecular docking, and MDS results are all in agreement and complementary. QSAR-VS successfully identified a more potent novel ARi and can be used in the development of therapeutic agents to treat diabetes.

1. Introduction

The aldose reductase (AR) belongs to the Aldo-Keto Reductase superfamily (AKRs) composed of 190 enzymes. AR catalyzes the reduction of carbonyl substrates such as sugar aldehydes along with few other biomolecules and marks their role in lipid, carbohydrates, and xenobiotic metabolism (Jez et al., 1997). AR is the first, rate-limiting enzyme in the polyol pathway and it causes sorbitol accumulation in the insulin-independent tissues. Hence, AR has been linked to retinopathy – cataract in particular – and the pathophysiology of diabetes sequelae such as angiopathy, neuropathy, and nephropathy (Moon et al., 2006).

AR inhibitors reduce diabetes-related disorders, particularly in the tissues that exhibit insulin-independent glucose uptake such as; neural tissues, the lens, and glomeruli. Various AR inhibitors such as, alrestatin, benurestat, epalrestat, fidarestat, imirestat, lidorestat have been developed to treat secondary complications in diabetes (Krans 1993, van Gerven and Tjon-A-Tsien 1995, Tsai and Burnakis 2016). ONO Pharmaceuticals developed Epalrestat, the only AR inhibitor approved for the treatment of diabetic neuropathies in Japan, India, and China (Steele et al., 1993, Kucerova-Chlupacova et al., 2020). Many of them have been studied in clinical trials, but eventually discontinued due to no effect or harmful side effects such as fever, nausea, diarrhea, increased liver enzymes, rash including toxic epidermal necrolysis, Stevens-Johnson syndrome, marked thrombocytopenia, lymphadenopathy, splenomegaly and adult dyspnea syndrome (Foppiano and Lombardo 1997). Due to these concerns, the need for the development of the de novo ARi inhibitor has been increased.

In QSAR study, statistical approaches are used to identify a mathematical correlation between structural properties of similar molecules and their bioactivity. Traditional QSAR study employs a variety of scientific disciplines including chemistry, computer science, mathematics, statistics, and biology. Following are the stages in a standard QSAR analysis procedure: (1) to Assemblage molecules with the specified activity/property (referred as a dataset); (2) to draw structures in 2D, convert them into 3D and to optimize them using an appropriate force field; (3) to calculate of as large as possible number of and variety of molecular descriptors and subsequent data pruning using a suitable statistical method; (4) to employ an appropriate feature (molecular descriptor) selection algorithm to build a QSAR model; and (5) adequate validation of the developed QSAR model (Pourbasheer et al., 2015, Masand et al., 2016).

To find out structural and molecular features of the molecule that govern the expected activity of the molecule is the main goal of a QSAR study (i.e. descriptive QSAR). Whereas prediction of dezired activity of the molecule prior to its wet lab synthesis and bio-testing is a secondary goal of the QSAR study (i.e. statistical QSAR) (Fujita and Winkler 2016). A good balance of descriptive and statistical elements in a QSAR model not only provides more information about the structural configurations that have a positive association with an intended activity/property of a drug candidate, but it also improves understanding of the drug's mechanism of action. To find a potential hit as an AR inhibitor, we have used QSAR, QSAR-based virtual screening, molecular docking, MD simulation and MMGBSA analysis in this study.

2. Materials and methods

2.1. Preparation of data sets

A crude dataset of 432 compounds with experimental AR inhibitory potency measured in terms of IC_{50} values is procured from ChEMBL (Gaulton et al., 2017) database. After removing structural duplicates, multi-component compounds or salts, and compounds with imprecise IC50 values finally a limited data set of 226 AR inhibitors with accurate IC_{50} values is obtained. IC_{50} values in nanomolar (nM) units were first converted into molar (M) and then into pIC_{50} M (pIC_{50} = -logIC_{50} M) for ease of data set handling. (See supplementary information **Table S1** for SMILES notations for all the 226 compounds with experimental IC_{50} and pIC_{50} M values).

2.2. Model development and validation

We did use ChemSketch to create 2D structures of all the 226 molecules, an Open Babel 3.1.1 programme (O'Boyle et al., 2011) to transfigure them into corresponding 3D structures, and a hyperchem programme (Ivanciuc 1996) to optimize these 3D molecular structures by employing the semi-empirical PM6 method (Bikadi and Hazai 2009). The resultant conformers were used for the calculation of the molecular descriptors using PyDescriptor available as a PyMOL plugin. More than 40,000 descriptors makeup the PyDescriptor, which covers topological, geometric, and constitutional chemical space of molecules (Masand and Rastija 2017).

Furthermore, the pre-filtration of descriptors (excluding semi-constant descriptors (greater than80%) and highly intercorrelated descriptors (greater than95%)) give rise to a contracted dataset of 2546 descriptors. The present data set of 226 molecules was divided into training (an 80% i.e. 181) for the QSAR model development and test sets (a 20 % i.e. 45) for its validation. The QSAR Models were built by multiple linear regression. The descriptor selection for the training set was carried out by using the entire subset and Genetic function algorithm available in QSARINS 2.2.4 software (Gramatica 2020) (Dearden et al., 2009, Gramatica 2013, Cherkasov et al., 2014, Gramatica 2014, Fujita and Winkler 2016, Harit et al., 2017).

To ascertain the robustness of the develop model various validation criteria are reported in the literature. The internal predictivity and statistical quality of the built model is verified by the parameters like the coefficient of determination (r^2), the

leave-one-out cross-validation (Q^2_{LOO}), and leave-many-out cross-validation (Q^2_{LMO}). Additionally, for every developed model the standard error of estimate (s) is described. To add an extra portion of the accuracy for the reported QSAR models, Root Mean Squared Error (RMSE) for the training ($RMSE_{tr}$) and external prediction ($RMSE_{ext}$) set that account on the whole error of the model are also described. To ensure least possible inter-correlation among descriptors, the QUICK rule was fixed to 0.05. A Y-randomization test at 2000 iterations was performed to ascertain the reliability of the developed QSAR model and to rule out the possibility of any speculative correlation.

The external validation of all the models were verified with the subsequent validation criteria; r^2_{ext} (external determination coefficient), Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, CCC_{ext}, r^2_m, and Δr^2_m. The parameter R^2m (overall) penalizes a model for large differences between observed and predicted values of the compounds of the whole set (considering both training and test sets). The Δr^2m estimates the indulgence between the values of the predicted and the resultant experimental activity data (pIC_{50} M value). It has been suggested that the observed value for the r^2_m should be greater than 0.5. All of the QSAR models were also tested for validation parameters, such as Golbraikh and Tropsha's criteria to justify model reliability and robustness.

Generally, good predictive ability of the developed QSAR model depends upon the closeness of predicted value against experimental biological activity value. Even, presence of a single outlier diminishes the predictive capability of the developed QSAR model. Subsequently, we have tried to highlight the outlier on the basis of those compounds who showed significantly high residual value in GA-MLR based QSAR models. Moreover, we have identified the outlier compounds by comparing the predicted value with the standardized residual values. Likewise, structural variation in database compounds was observed by leverage effect in Williams plot. The applicability domain of the developed QSAR model is ascertained by merging the leverage and the standard residuals.

The multi criteria decision making (MCDM) function available in QSARINS v4.1.1 software is used to rank the developed models. It includes some criteria related to the external and internal validation whose values fall between 0 and 1, wherein 0 represents worst validation while 1 represents the best validation performance. The geometric average of all the values obtained in internal and external validation gives rise to MCDM values. The model with the high MCDM agreement among the best selected validation criteria is sorted as the best QSAR model for the analysis. In addition, the best model is evaluated for OECD (organization for economic corporation and development) guidelines (Gramatica et al., 2013, Gramatica et al., 2014, Consonni et al., 2019).

2.3. QSAR based virtual screening

In QSAR based virtual screening, we have carried out scaffold hopping by using the RDKIT module. Herein, we have used the most active compound 18 as a template molecule to generate different variants. This has given rise to 14 different scaffolds with enhanced chemical space. Accordingly, 14 variants of compound 18 were used for QSAR-based virtual screening. Erstwhile to molecular descriptor calculations, the 3D- structures of the molecules were arranged in the same way as a modeling set. Then molecular descriptors were calculated and the appropriately validated five parametric QSAR model was used to envisage the biological property of novel compounds (Jawarkar et al., 2021). The structures of some representative scaffolds are given in Fig. 1.

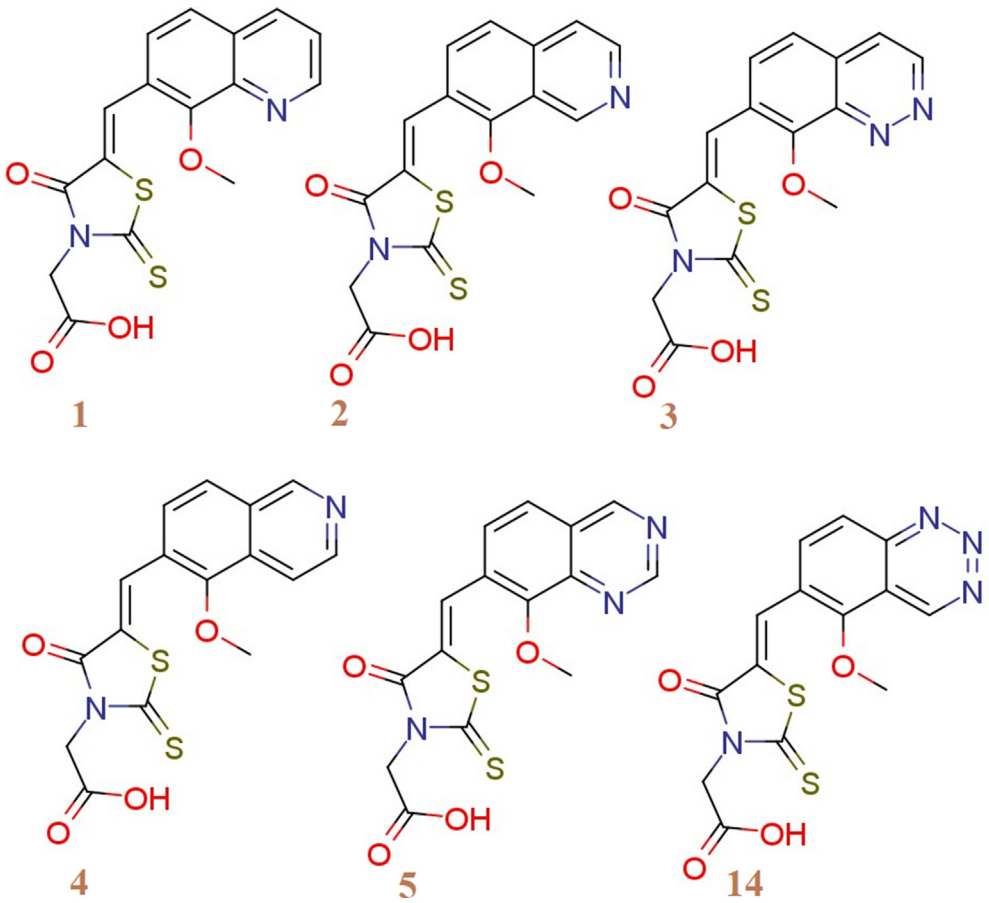


Fig. 1. Showing few representative examples of compounds generated by scaffold hopping.

(Depiction of 14 scaffolds, their smiles notations, calculated descriptors and predicted IC_{50} Values is available as **Table S2** in supplementary material).

2.4. Molecular docking study experimental

The human AR in the pdb file was obtained from the Protein Data Bank (https://www.rcsb.org/structure/1fzd). The pdb:1fzd was selected on the origin of X-ray resolution and conclusion of the sequence. The optimized protein is suitable for docking analysis. All the active compounds were docked in the active site, but for the sake of ease, herein, the docking pose for most active molecule 18 as a symbol has been depicted.

The software NRGSuite was used to investigate molecular docking. This open source programme is available as a PyMOL plugin (https://www.pymol.org). It has the competence to ascertain the surface cavities in a protein and use them as target binding-sites for docking simulations with the aid of FlexAID. It practices genetic algorithms for conformational search, simulates ligand and side-chain flexibility and permits for the simulation of covalent docking. In the current work, flexible-rigid docking procedure was hired with succeeding default settings to get finest performance from NRGSuite: binding sites input method, spherical shape; spacing of three dimensional grid: 0.375 Å; side chain flexibility- no; ligand flexibility- yes; ligand pose as reference- no; constraints- no; HET groups- comprised water molecules; van der Waals permeability-0.1; solvent types- no type; number of chromosomes- 1000; number of generations- 1000; fitness model- share; reproduction model- population boom; number of TOP complexes-5, Biovia Discovery studio software was used to visualize the docking results (Gaudreault et al., 2015).

2.5. MD simulations study

Based on the virtual screening results, the scaffold 5 with a docking score of −8.08 kcal/mol and Molecule 18 (-7.91 kcal/mol) is further investigated in molecular dynamics and simulation using the Schrodinger Desmond module (MD simulation) and MMGBSA binding free energy analysis. The protein ligand docking complexes of scaffold 5 and molecule 18 docking complexes were created using the Desmond module's SPC (Simple point charge) configuration. The OPLS-2005 force field (Bowers et al., 2006) and explicit solvent model with the SPC water molecules were used in this system (Jorgensen et al., 1996, Shivakumar et al., 2010). Na^+ ions were added to neutralize the charge. 0.15 M, NaCl solutions added to the system to simulate the physiological environment. The NPT ensemble was set up by using the Nose-Hoover chain coupling scheme (Martyna et al., 1992) with temperature 300 K, relaxation time of 1.0 ps and pressure 1 bar was maintained in all the simulations. A time step of 2 fs was used. The Martyna-Tuckerman–Klein chain coupling scheme (Martyna et al., 1994)barostat method was used for pressure control with a relaxation time of 2 ps. The particle mesh Ewald method (Toukmaji and Board 1996) was used for calculating long-range electrostatic interactions and the radius for the Coulomb interactions was fixed at 9 Å. RESPA integrator was used to calculate the non-bonded forces. The root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), protein ligand interactions were monitored to observe the stability of the complex in MD simulations.

2.6. Molecular mechanics generalized born and surface area (MMGBSA) calculations

During MD simulations of aldose reductase enzyme in complexed with molecule 18 and scaffold 5, the binding free energy (Gbind) of docked complexes was calculated using the premier molecular mechanics generalized Born surface area (MM-GBSA) module (Schrodinger suite, LLC, New York, NY, 2017–4). The binding free energy was calculated using the OPLS 2005 force field, VSGB solvent model, and rotamer search methods (Jawarkar et al., 2022, Kumar et al., 2022). After the MD run, 10 ns intervals were used to choose the MD trajectories frames. The total free energy binding was calculated using equation (1):

$$\Delta Gbind = Gcomplex - (Gprotein + Gligand) \quad (1)$$

where, ΔGbind = binding free energy, Gcomplex = free energy of the complex, Gprotein = free energy of the target protein, and Gligand = free energy of the ligand. The MMGBSA outcome trajectories were analyzed further for post dynamics structure modifications.

2.7. In silico ADMET property prediction

The ADME (Absorption, Distribution, Metabolism, and Excretion) properties of the chosen phytocompound were calculated using the SWISS adme web server (https://www.swissadme.ch/). The molecular as well as pharmakokinetic behaviour of molecule 18 and scaffold 5 were assesed for drug like properties, total polar surface area (TPSA), water solubility, blood brain barrier (BBB) permeant, gastro intestinal (GI) absorption, Lipinski, Ghose, Veber, Egan and Muegge violations and synthetic accessibility. If the test molecules e.g. molecule 18 and scaffold 5 pass the drug like filters therefore, it can be predicted to be candidate molecule against aldose reductase. For the analysis of ADME, the SMILEs format of molecule 18 and scaffold 5 have been used individually to retrieved the outcomes from the server.

3. Results

In the present investigation, QSAR analysis is performed using a dataset composed of 226 ARi inhibitors with reported IC_{50} values and molecular docking is accomplished to pinpoint the important structural features. The QSAR model was developed by using easily interpretable molecular descriptors to correlate them with structural features. The five parametric GA–MLR based QSAR model has decent external predictive ability with the incidence of easily comprehensible molecular descriptors alongside the interpretation in terms of structural features. For the aim of model selection, Multi-Criteria Decision Making (MCDM) analysis has been employed in this study. As a result of the MCDM study, a robust QSAR model was chosen for the analysis. The findings are given below.

3.1. Multi-Criteria decision making (MCDM) analysis

In the current QSAR analysis, we have implemented Multi-Criteria Decision Making (MCDM) (Pavan and Todeschini 2009) technique that involves combining the performances of a certain number of criteria simultaneously, as a single number (score) between 0 and 1.

The MCDM is achieved by associating to each validation criteria a desirability function whose value falls in the range from 0 to 1 (where 0 designates the worst validation criteria value and 1 the best). Using the geometric average of all the values obtained from the desirability functions gives the MCDM value. By default, the MCDM of fitting (maximizing R^2, R^2_{adj} and CCC_{tr}, while minimizing R^2-R^2_{adj}), cross validation (maximizing Q^2_{LOO}, Q^2_{LMO} and CCC_{cv}, while minimizing R^2_{Yscr}) and external validation (maximizing Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{EXT}), are automatically calculated using all the above criteria in QSARINS v2.2.4 programme. If any one of the criteria is missing, then the MCDM model will not be obtained. The model with the best

Table 1
Display of five variable models in MCDM.

Model id	Size	Variables/Descriptor
134	5	com_ringCminus_2A H_ringN_2B com_Nhyd_2A notringC_don_3B allminus_SASA
133	5	H_ringN_2B lipo_don_3Bc com_Nhyd_2A notringC_don_5B allminus_SASA
132	5	H_ringN_2B lipo_don_3Bc com_Nhyd_2A notringC_don_3B allminus_SASA
129	5	minus_H_3B H_ringN_2B com_Nhyd_2A notringC_don_3B allminus_SASA
131	5	all_HASA3 H_ringN_2B com_Nhyd_2A notringC_don_3B allminus_SASA
130	5	all_HASA3 H_ringN_2B com_Nhyd_2A notringC_don_5B allminus_SASA
128	5	H_ringN_2B ringN_H_8Ac com_Nhyd_2A notringC_don_5B allminus_SASA
127	5	H_ringN_2B ringN_H_8Ac com_Nhyd_2A notringC_don_3B allminus_SASA
126	5	H_ringN_2B ringN_H_5Ac com_Nhyd_2A notringC_don_5B allminus_SASA

MCDM criteria is selected for analysis(**See** Table 1). In the present analysis, model no 134 has satisfied the best MCDM criteria (**See** Table 2 **for various MCDM parameters**), therefore it is selected for the analyses (see Fig. 2).

In Fig. 3, model no 135 depicts very good fitting performance but it is lacking in external predictive performance. For instance, model no 133 showed very good external predictive performance but not satisfactory in fitting. Therefore, model no 134 in between 133 and 135 showed better compromise between fitting and predictivity, hence selected as best model for analysis and QSAR based virtual screening. Moreover, apart from model no 134, we have developed a full set model.

3.2. QSAR model 1.1 (Model no 134)

pIC_{50} = 2.603 (±0.447) − 0.193 (±0.085) **com_ringCminus_2A** + 0.249 (±0.074) **H_ringN_2B** − 1.331 (±0.226) **com_Nhyd_2A** − 0.34 (±0.113) **notringC_don_3B** + 0.012 (±0.001) **allminus_SASA**.

(The experimental and predicted pIC_{50} M value for the divided dataset model is available as **Table no S3** in supplementary material).

R_2:0.80, R^2_{adj}:0.79, R^2-R^2_{adj}:0.01, LOF:0.26, K_{xx}:0.20, DeltaK:0.09, $RMSE_{tr}$: 0.49, MATer:0.40, RSS_{tr}:42.75, CCC_{tr}:0.89, s:0.50, F:140.68, Q^2_{loo}:0.79, R^2-Q^2_{loo}:0.02, $RMSE_{cv}$: 0.50, MAE_{cv}: 0.41, $PRESS_{cv}$: 45.25, CCC_{cv}: 0.88, Q^2_{LMO}: 0.79,R^2Y_{scr}: 0.03, Q2 scr: −0.04, RMSE AVY$_{scr}$: 1.07, $RMSE_{ext}$: 0.54, MAE_{ext}: 0.44, $PRESS_{ext}$:12.98, Q^2_{F1}: 0.80,Q^2_{F2}: 79, Q^2_{F3}: 0.75,CCC exe: 0.88, r^2m aver.: 0.68, r^2m delta: 0.18.

3.3. QSAR model 1.2 (Full set model)

$_{P}IC_{50}$ = 2.619 (±0.399) − 0.085 (±0.033) **minus_don_3B** + 0.431 (±0.086) **don_ringC_6Ac** − 0.353 (±0.166) **don_notringC_4B** − 0.584 (±0.123) **fsp2Osp2C4B** + 0.012 (±0.001) **allminus_SASA** (The experimental and predicted $_{pIC50\ M}$ value for the full dataset model is available as table no S4 in supplementary material).

R^2 = 0.79, R^2_{adj} = 0.79, R^2-R^2_{adj} = 0.01, LOF = 0.28, K_{xx} = 0.16, DeltaK = 0.14, $RMSE_{tr}$ = 0.50, MAE_{tr} = 0.40, RSS_{tr} = 57.41, CCC_{tr} = 0.89, s = 0.51, F = 168.60, Q^2_{loo} = 0.78, R^2-Q^2_{loo} = 0.01, $RMSE_{cv}$ = 0.52, MAE_{cv} = 0.41, $PRESS_{cv}$ = 60.39, CCC_{cv} = 0.88, Q^2_{LMO} = 0.78, R^2Y_{scr} = 0.02, Q^2Y_{scr} = -0.03, RMSE AV Y_{scr} = 1.10.

The statistical validation parameters listed above are recommended for judging internal and external robustness, and they have the same meaning as before (see Supplementary Material Table S5 for detailed formulae and Table S6 for the detailed descriptions of the descriptors). The high value of R^2_{tr} (coefficient of determination), R^2_{adj}. (adjusted coefficient of determination), R^2_{cv} (Q^2_{loo}) (cross-validated coefficient of determination for leave-one-out), R^2_{ex} (external coefficient of determination), Q^2_{Fn}, and CCC_{ex} (concordance correlation coefficient), and the low value of LOF (lack-of-fit), $RMSE_{tr}$ (root mean square error).The various graphs associated with the model(**see** Figs. 3 and 5), such as MAE_{tr} (mean absolute error), R^2_{Yscr} (R^2 for Y-scrambling), and others, show that the model is statistically robust, with high internal and external predictive capacity, and is free of chancy correlation. Furthermore, the Williams plot (**see** Fig. 4) demonstrates that the model is statistically acceptable.Moreover, insubris As a result, it complies with all of the OECD's suggested standards for developing an effective QSAR model.

4. Discussion

4.1. Allminus_SASA

The solvent accessible surface area, integrates shape and electrical property which additionally can be contributed by the presence of hydrogen bond donor/acceptor atoms (heteroatoms, such as N, O, etc.) in a molecule and is determined by mapping atomic partial charges. Owing to the electronegativity difference between such heteroatoms and carbon/hydrogens there occurs a partial charge separation, subsequent dipole formation, thereby intensifying the drug – water solubility and facilitating drug – biological target interaction, ultimately. The molecular descriptor, allminus_SASA (solvent accessible surface area ($Å^2$) of all negatively charged atoms) is in a positive relationship with pIC_{50} M (positive coefficient in both QSAR equation) and hence, increase in its value may possibly lead to better AR inhibitory activity.

This can be well illustrated by comparing compound 190 (allminus_SASA = 131.34, Number hydrogen bond donor/acceptor = 2; pIC_{50} = 4.0 M) and 16 (allminus_SASA = 411.21,Number hydrogen bond donor/acceptor = 7; pIC_{50} = 7.48 M) wherein possibly, for about three fold increase in allminus_SASA value leads to more than 3000 fold increase in AR inhibitory potency. Illustrative Fig. 6 depicts the allminus_SASA.

4.2. Minus_don_3B

The minus_don_3B (presence of a donor within three bonds from a negatively charged atom) has a negative relationship with the inhibitory activity of AR and its value should be kept as low as possible. The importance of this bioactivity governing feature can be supported by comparing compound 20 (minus_don_3B = 0; pIC_{50} = 6.40 M) with the compound 172 (minus_don_3B = 12; pIC_{50} = 4.69 M). Both, compounds 20 and 172 constitute a comparable number of hydrogen bond donors/acceptors which are expected to bring very close AR inhibitory potency to both (as explained for allminus_SASA in preceding section). But contrast observation in case of the compound 20 and 172 pair, highlights the importance of absence of not all but specific donor atoms or negatively charged carbons, defined by minus_don_3B. Furthermore, the lone pair of electrons, associated with the donor NH & OH groups can be deployed to make a covalent link with a biological target. Plausibly, a good compromise between phenyl and moderately long chain ether functionality induced lipophilicity and donor atoms induced hydrophilicity, enhanced the AR inhibitory potency of compound 20 (**see** Fig. 7).

4.3. Don_ringC_6Ac

The Don_ringC_6Ac (presence of partially charged ring carbon atoms within 6 Å of the donor) being having positive correlation

Table 2
Presentation of different MCDM parameters.

Model id	R^2	R^2_{adj}	Q^2_{LOO}	Q^2_{LMO}	CCC_{tr}	CCC_{cv}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}
134	0.7999	0.7942	0.7881	0.7888	0.8888	0.8824	0.7975	0.7935	0.7488
133	0.7984	0.7926	0.7848	0.7821	0.8879	0.8805	0.8164	0.8128	0.7724
132	0.7965	0.7908	0.7838	0.7803	0.8867	0.8798	0.8163	0.8127	0.7722
129	0.7953	0.7895	0.7816	0.7782	0.886	0.8785	0.7435	0.7385	0.682
131	0.7951	0.7893	0.7833	0.7814	0.8858	0.8794	0.7945	0.7904	0.7451
130	0.7947	0.7889	0.7817	0.7766	0.8856	0.8785	0.7882	0.784	0.7373
128	0.7942	0.7883	0.7802	0.7773	0.8853	0.8777	0.7859	0.7817	0.7345
127	0.7933	0.7874	0.78	0.7767	0.8847	0.8775	0.7873	0.7831	0.7362
126	0.7927	0.7869	0.7782	0.7759	0.8844	0.8765	0.7814	0.7771	0.729

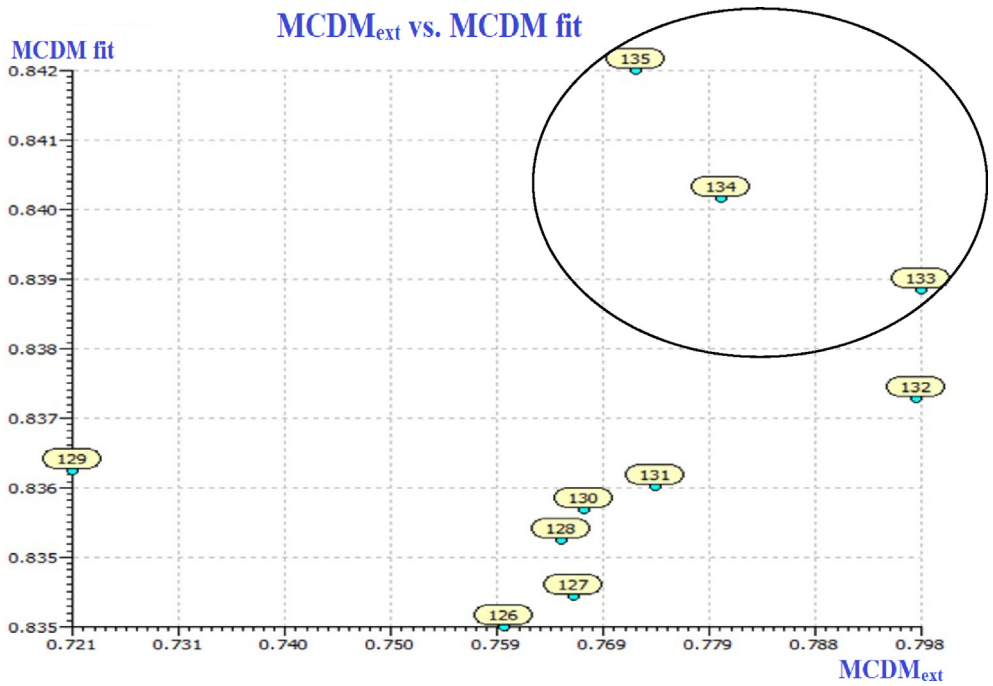


Fig. 2. Depiction of MCDM graph for model 135. The × axis indicates MCDM fit while y axis shows MCDM ext (Black circle showing three best models).

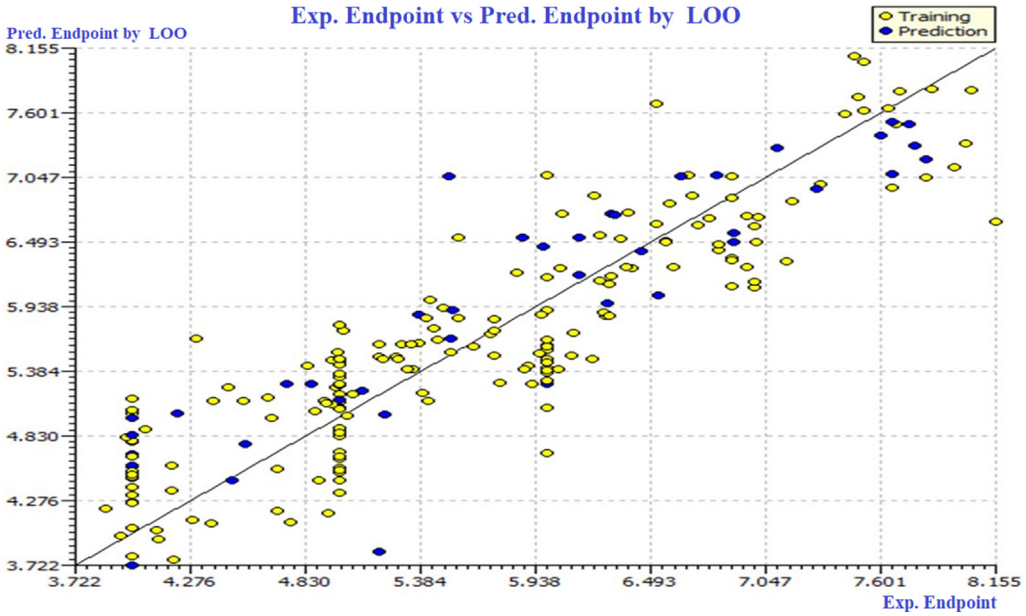


Fig. 3. Depiction of Scatter plot of experimental vs. predicted data by LOO.

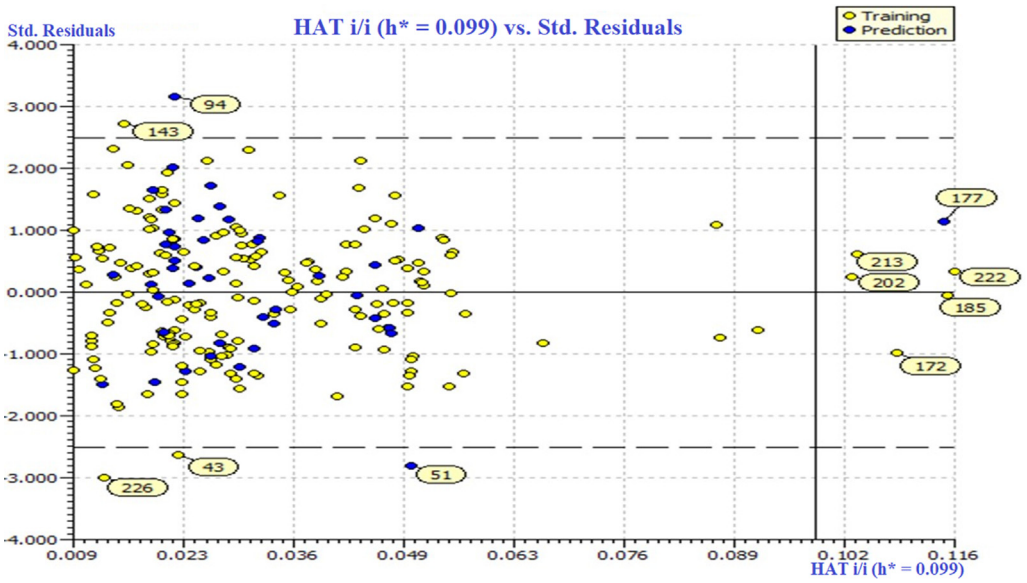


Fig. 4. Display of Williams plot using data predicted by LOO. (Molecules out of applicability domain have been shown with their serial numbers).

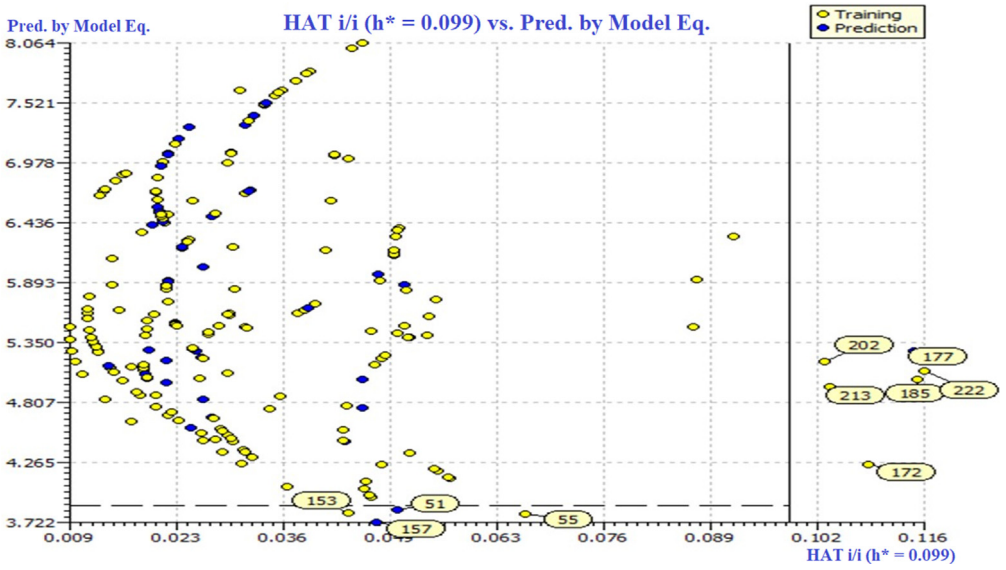
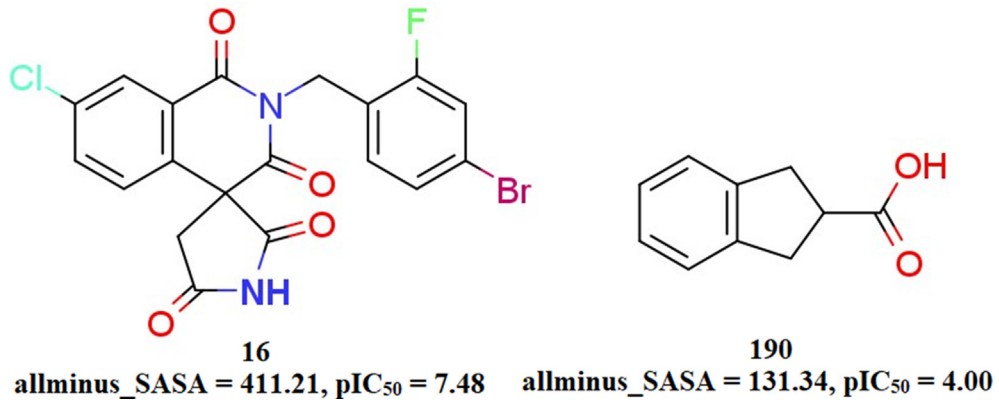


Fig. 5. Insubria plot for Divided set model 134. (Molecules out of applicability domain have been shown with their serial numbers).



16
allminus_SASA = 411.21, pIC$_{50}$ = 7.48

190
allminus_SASA = 131.34, pIC$_{50}$ = 4.00

Fig. 6. Depiction of the descriptor allminus_SASA for compound 16 & 190. (The negatively charged atoms are depicted by a bold red color).

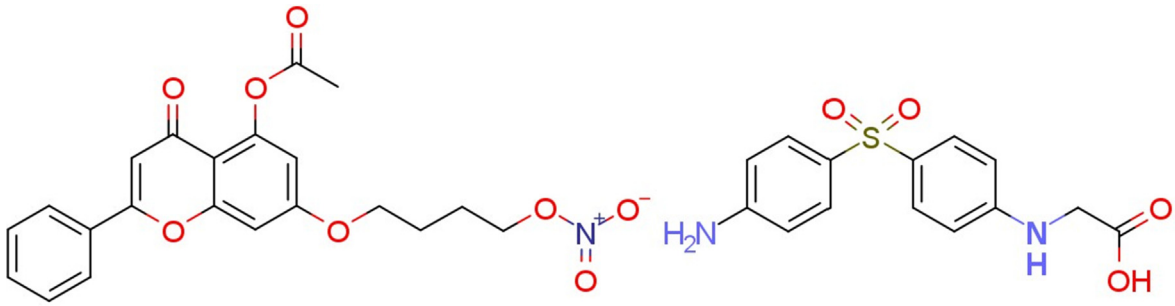


Fig. 7. Depiction of the descriptor minus_don_3B for the compound 20 and 172 only. (Donors are indicated by the dotted circle while negatively charged atoms are shown by red bold color).

with the pIC_{50} (M), high possible value of don_ringC_6Ac is advisable for better AR inhibitor lead optimization. In Compound 85 (don_ringC_6Ac = 2.49, number of H-bond donor/acceptor = 9; pIC_{50} = 7.52 M), a blend of hydrogen bond donors capable of forming electrostatic interactions, with equally important lipophilic carbons (in unsaturated region of the isoquinoline and the pyrrolidine ring) enabling lipophilic interactions of molecule, for AR inhibition makes it 2.83 unit more potent than compound 103 (don_ringC_6Ac = 0, number of H-bond donor/acceptor = 7; pIC_{50} = 4.59 M) (**see** Fig. 8).

4.4. Don_notringC_4B

The don_notringC_4B (non-ring carbon atom within four bonds from the donor) has negative correlation with biological activity and hence minimum possible value of this molecular descriptor while AR inhibitor lead optimization is recommended. Absence of non– ring carbon atoms within four bonds from the donor atom probably makes compound 46 about 660 times more potent than compound 185 where there are two such non-ring carbons present (Fig. 9). In corollary, a non-ring carbon atom within a molecule must be placed at a distance of five or more bonds from the donor atoms to develop leads with desired potency.

4.5. fsp2Osp2C4B

A molecular descriptor fsp2Osp2C4B (the frequency of occurrence of a sp^2 hybridized carbon atom exactly four bonds from a sp2 hybridized oxygen atom) has a negative correlation with biological activity and keeping its value minimum possible is advisable for better AR inhibitory potency. Absence of this specific combination of sp2-O a distance of four bonds from the sp2-C atom in compound 37 possibly make it more potent than compound 1 in which such a combination is observed twice. (Fig. 10). Furthermore, overall lipophilicity of the molecule to the large extent is contributed by unsaturated ring carbons. Importance of lipophilicity in molecules for its AR inhibitory potency highlighted in foregone discussion support and supported by the importance of absence of not all but specific combination of sp2C and sp2O.

4.6. com_ringCminus_2A

com_ringCminus_2A encodes information on the number of negatively charged ring Carbon atoms within 2 Å from the center of mass of the molecule. AR inhibitory potency is inversely proportional to the value of this molecular feature and hence, low possible value of **com_ringCminus_2A in a molecule** is advisable for better AR inhibitory potency. This observation can be supported by Comparison of 203 (**com_ringCminus_2A** = 4; pIC_{50} = 4.000 M) with 109 (**com_ringCminus_2A** = 0; pIC_{50} = 7.95 M) wherein decrease in value of **com_ringCminus_2A f**rom 4 in compound 203 to 0 in compound 109 cause about 10,000 fold increase in AR inhibitory activity (see Fig. 11).

Generally, molecules in the receptor pocket should be more steadily balanced if its center of mass is below the balance point, but unstable if it is above the balance point. The appropriate orientation of the molecule within the active site pocket is determined by the molecule's center of mass. This finding supports the idea that having a larger number of negatively charged carbon atoms within the 2 Å radius can cause problems with molecular alignment within the receptor active site pocket. This could be the cause

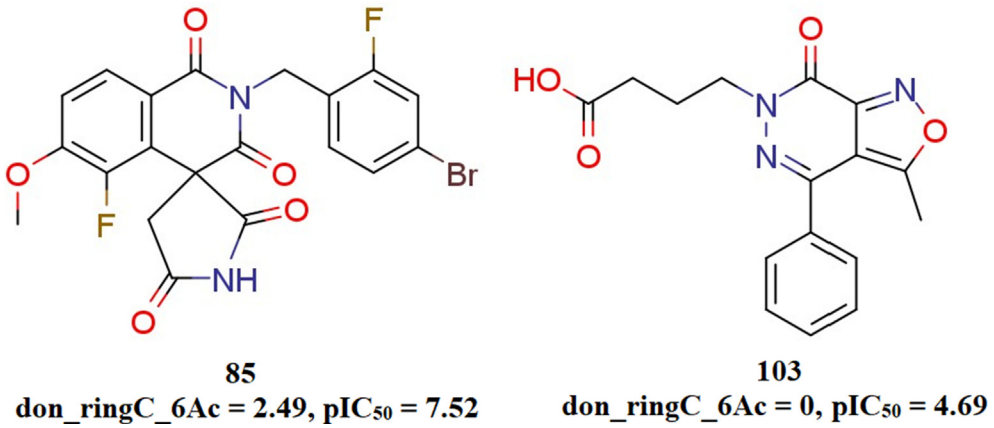


Fig. 8. Representation of descriptor Don_ringC_6Ac for the compound 85 and 103 only. (Donors are highlighted by a dotted circle while carbon atoms within 6 A^0 are indicated by red bold color).

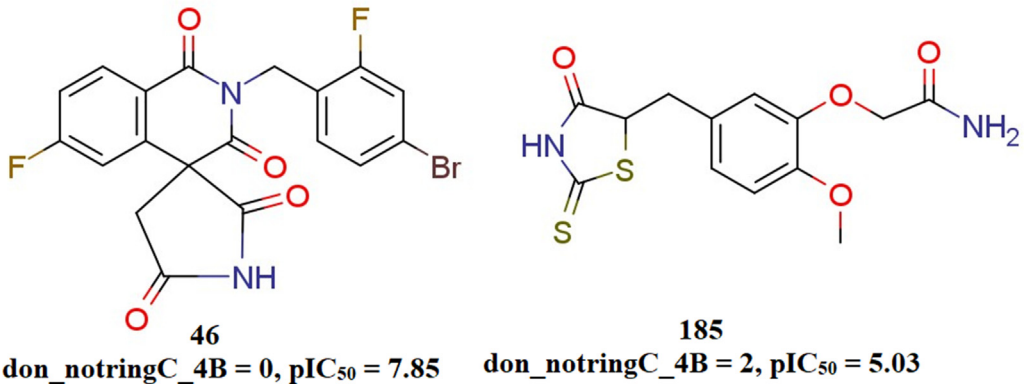


Fig. 9. Depiction of descriptor don_notringC_4B for the compound 46 and 185.

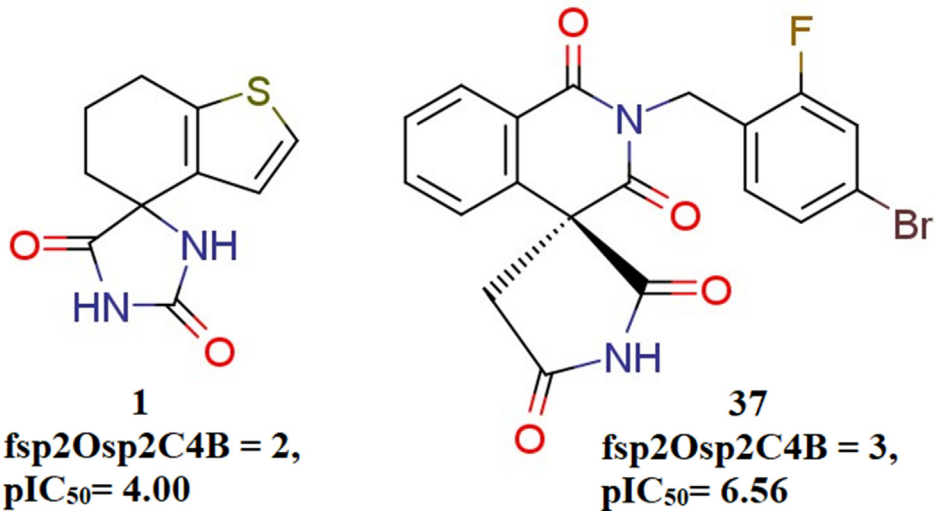


Fig. 10. pictorial depiction of descriptor fsp2Osp2C4B for the compounds 1 and 37.

of the differences in biological activity between molecules 203 and 109 (**see** Fig. 11).

4.7. com_Nhyd_2A

In the established QSAR model, the molecular descriptor com_Nhyd_2A (presence of hydrophobic nitrogen within 2 Å of the center of mass) is negatively connected with biological activity and hence it is advisable to abstain from introduction of such hydrophobic nitrogen (partial charge in between −0.2 to + 0.2) to avoid deterioration of biological activity. Here, in compound 57, one hydrophobic nitrogen is present with a charge of −0.09 while another nitrogen has a charge of 0.06. (**see** Fig. 12).

The poor activity for the molecule 16, 22, 36, 26, 32, 33 and 51 can be correlated with the high value of such hydrophobic Nitrogens within $2A^0$ from the center of mass of the molecule (com_Nhyd_2A = 1).

4.8. H_ringN_2B

The descriptor H_ringN_2B represents a combination of ring nitrogen and hydrogen separated by within 2 bonds. This descriptor has a positive coefficient in the QSAR model; therefore the combination of ring nitrogen within 2 bonds from the hydrogen element is the favorable combination to be used for hit/lead optimization. (**see** Fig. 13).

Generally, Hydrogen is the smallest element, it implies that there should be minimum bulk in the vicinity of ring Nitrogen atoms. Therefore, in future modifications, steric bulk nearer to ring Nitrogen atoms should be avoided to have better activity against AR.

A poor activity profile of molecules 112, 113, 100, 102, 118 and 133 is due to absence of such hydrogen and ring nitrogen combinations within 2 bonds (H_ringN_2B = 0), therefore higher the number of such combinations, greater will be AR inhibitory activity. This may be a plausible reason for the difference in the biological activity of these molecules.

4.9. QSAR based virtual screening

The QSAR based virtual screening was performed by employing a divided set QSAR model.

Among these 14 scffolds furnished by QSAR-VS, scaffold 5 turned out as the more potent AR inhibitor than a template compound 18. Close observation of the values of molecular features for both scaffold-5 and compound 18 revealed that the large value of two molecular descriptors viz. number of Hydrogen atoms that are two bond away from ring Nitrogen atoms (*H_ringN_2B*) and solent accessible surface area of all the negatively charged atoms

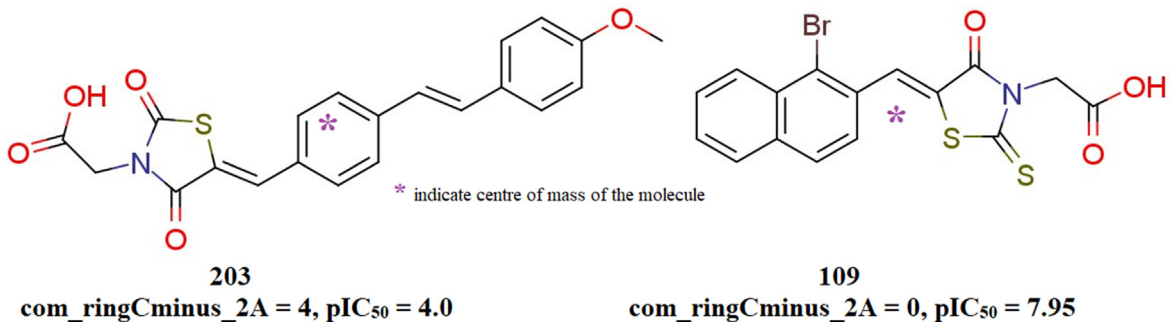


Fig. 11. Presentation of the descriptor com_ringCminus_2A for the compound 203 and 109.

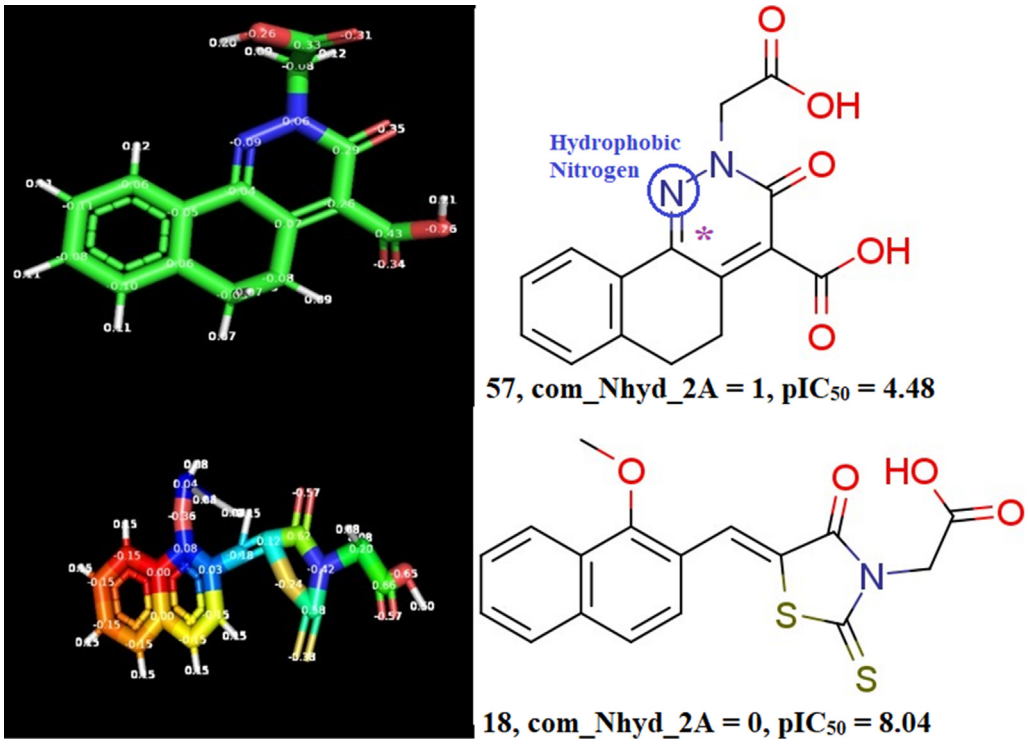


Fig. 12. Representation of the descriptor com_Nhyd_2A for the compound 57 and 18(Star in the figure indicates center of mass of the molecules).

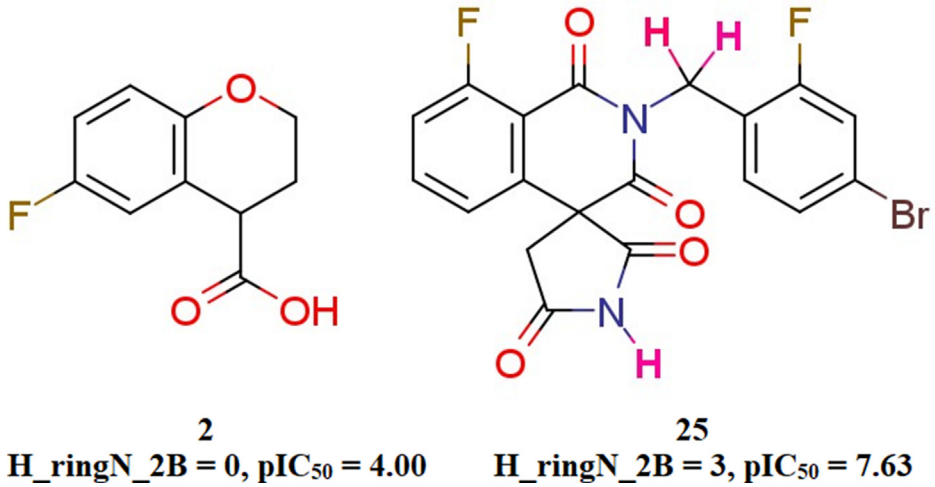


Fig. 13. Depiction of the descriptor H_ringN_2B for the compound 2 and 25.

(allminus_SASA) which are in positive corelation with AR inhibitory potency suffice to enhance the AR inhibitory potency of the scaffold-5 (*H_ringN_2B* = 4, allminus_SASA = 353.8; pIC_{50} = 8.05 M) over template compound-18 (H_ringN_2B = 2, allminus_SASA = 3 37.4; pIC_{50} = 8.04 M). In other scaffolds, increase in either *H_ringN_2B* or allminus_SASA or both is not as in case of scaffold −5. Hence, the synchronous effect of *H_ringN_2B* and allminus_SASA is the possible reason for the increase in the potency of AR inhibitor (scaffold 5).

4.10. Molecular docking analysis of molecule 18

Human AR (Pdb-1fzd), a 36 kDa enzyme with a supple and malleable active site, is the target protein under consideration. The enzyme folds into a TIM-barrel and is thought to be involved in diabetic retinopathy and angiopathy. (Docking score, RMSD and pedicted $^{pIC}50\ ^{M}$ values are displayed in Table 3).

As a result, it was created to be a promising drug target. It uses NADPH as a reducing cofactor to convert various aldehydes (including glucose in diabetic circumstances) to their corresponding alcohols. NADPH contributes a hydride ion to the carbonyl carbon of the aldehydes, resulting in a negatively charged intermediate, although the exact mechanism is still debated. Furthermore, a subsequent proton transfer from one of the nearby acidic active site residues is used to trace this step. The binding site consists of two sub-pockets, one encompassing the residues possibly involved in catalysis (Tyr48, Lys77, and His110) sideways with the nicotinamide moiety of the cofactor, while the second so-called specificity pocket is formed by Trp111, Ala299, Leu300 and Phe122.

With a docking score of −7.912 kcal/mol, molecule 18 orients in the specificity pocket in the same way as the pdb-1fzd ligand (**see** Fig. 14). Through typical hydrogen bonding, carbon hydrogen bonding, and hydrophobic interactions, molecule 18 forms a complex with human AR. Molecule 18 resides in the second pocket, known as the specificity pocket, in which the 4-oxothiazolidine oxygen acts as an acceptor, forming a conventional hydrogen bond with the NH2 hydrogen atom of specificity pocket residue A: LEU300 (1.90 A0) and so acting as a hydrogen bond donor. Further the same 4-oxo thiazolidine oxygen acceptor atom anchored carbon hydrogen bond with hydrogen atom of ALA299 (2.74A0) residue therefore, acts as hydrogen bond donor in the drug receptor interaction. (**see** Fig. 15).

Furthermore, molecule 18′s terminal acceptor carboxyl oxygen forms a carbon hydrogen bond with the specificity pocket A:

Table 3
Depiction of docking Results along with pIC_{50} M and IC_{50} by QSAR based virtual screening for the Molecule 18 and the series of 14 Scaffolds.

Sn	DockingScore (kcal/mol)	RMSD	pIC_{50} M	IC_{50} in nM
Molecule 18	**−7.91**	**0.95659**	**8.04**	**9.12**
Scaffold 1	−7.89	1.02	6.21	616.5
Scaffold 2	−7.69	1.28	7.83	14.79
Scaffold 3	−7.86	1.41	7.61	24.54
Scaffold 4	−7.81	1.65	7.83	14.79
Scaffold 5(Pose 1)	**−8.08**	**1.29**	**8.05**	**8.91**
Scaffold 5(Pose 2)	**−8.00**	**0.88**		
Scaffold 6	−7.91	1.20	7.84	14.4
Scaffold 7 (Pose 1)	−8.06	1.11	7.83	14.7
Scaffold 7(Pose 2)	−8.04	1.25		
Scaffold 8	−7.87	1.82	7.52	30.2
Scaffold 9	−7.91	1.44	7.79	16.2
Scaffold 10	−7.83	1.18	7.97	10.7
Scaffold 11	−7.91	0.81	7.74	18.1
Scaffold 12	−7.96	1.50	7.61	24.5
Scaffold 13	−8.01	1.21	7.75	17.7
Scaffold 14	−8.00	0.99	7.83	14.7

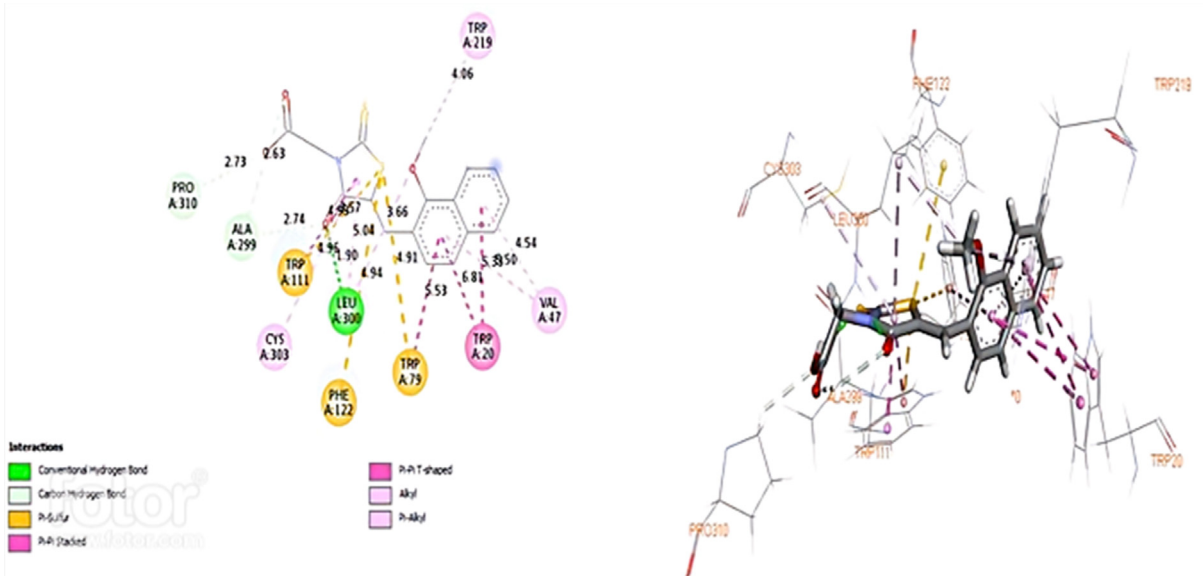


Fig. 14. Showing 3D and 2D interaction of Molecule 18 with Human aldose Reductase.

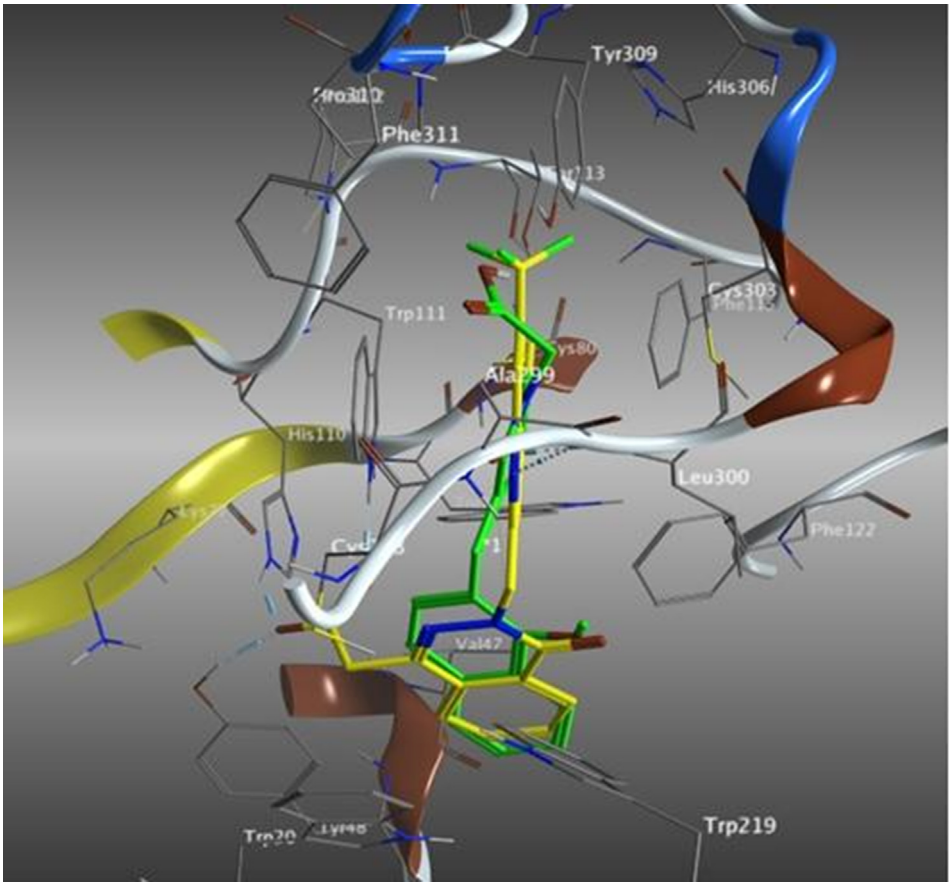


Fig. 15. Depiction of superimposed orientation of Molecule 18 (green) with Pdb-1fzd ligand (yellow) in specificity pocket.

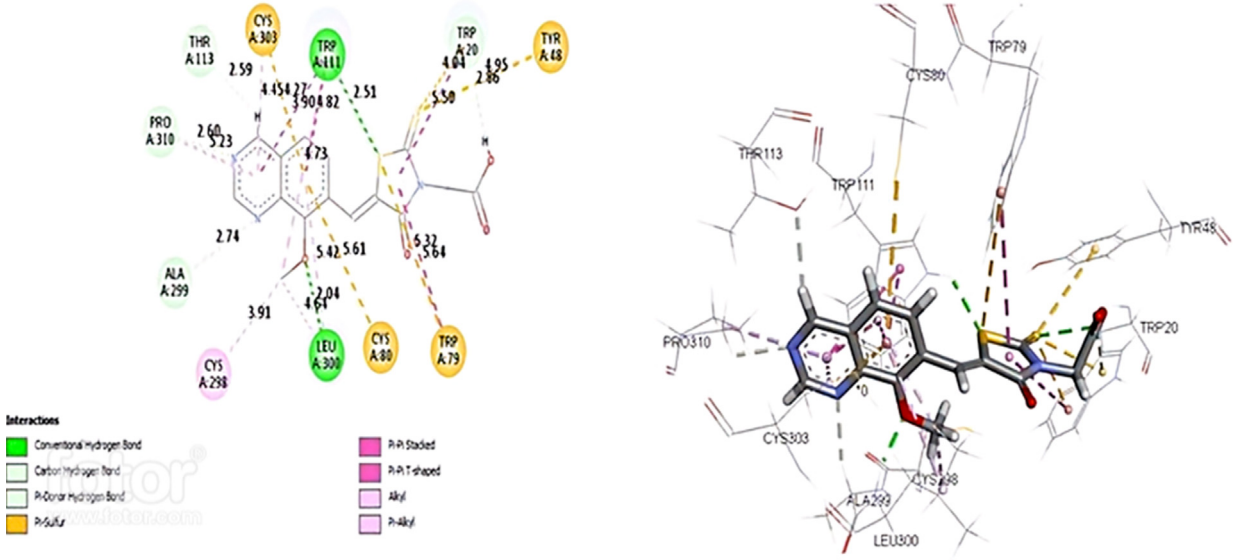


Fig. 16. Showing 3D and 2D interaction of Scaffold 5 with Human aldose Reductase.

ALA299 (2.62A0) residue, where the side chain carbon atom acts as a donor in drug receptor interactions. Furthermore, the acceptor hydroxy group of the terminal carboxyl substituent forms a carbon hydrogen bond with the donor hydrogen atom of the secondary amino group of PRO310 (2.74 Å) residue, producing the human aldose reductase specificity pocket. Following that, the naphthalene component of Molecule 18 intercalates with the benzene ring of TRP 20(5.21 Å) residue, resulting in a π - π stacking hydrophobic interaction between the naphthalene ring and the benzene ring of TRP20 residue due to the involvement of π orbitals.

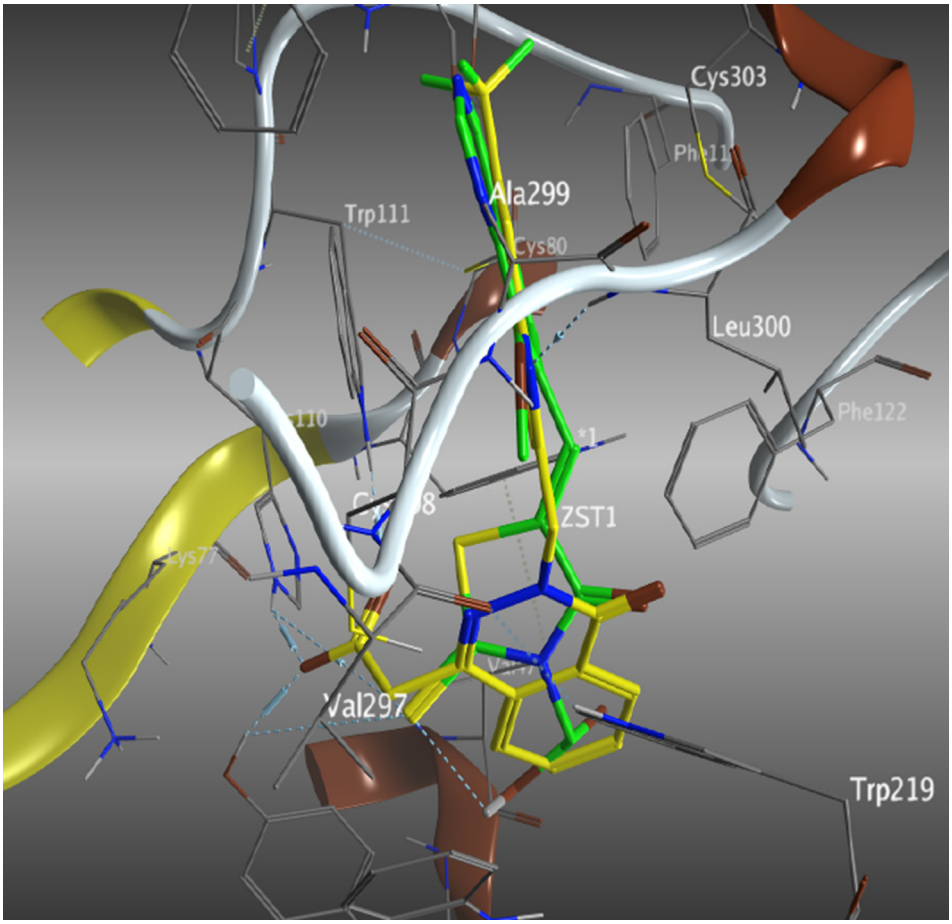


Fig. 17. Depiction of superimposed orientation of scaffold 5 (green) with Pdb-1fzd ligand (yellow) in specificity pocket.

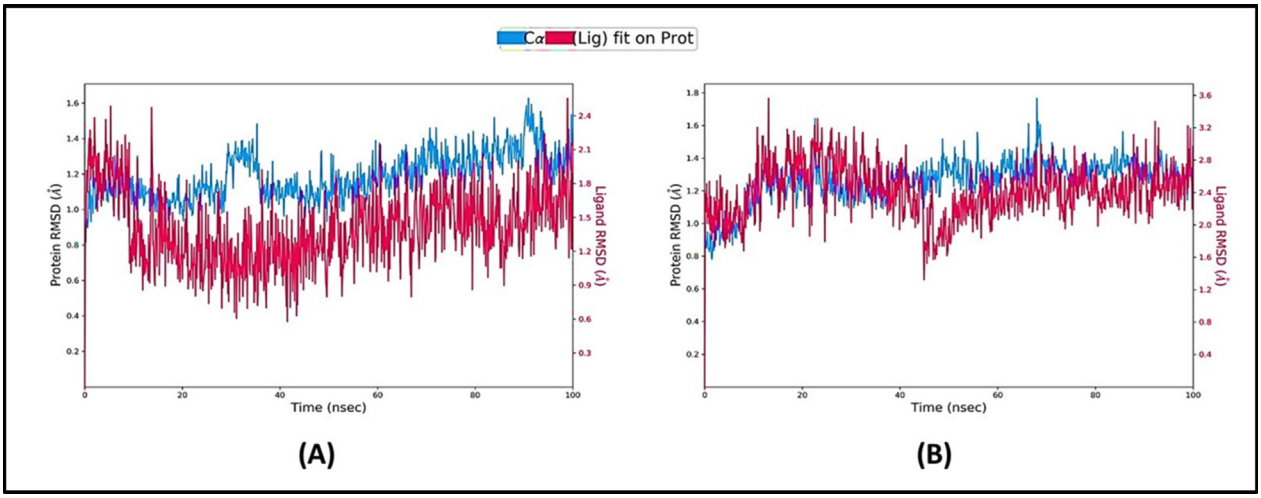


Fig. 18. Presentation of human AR with (A) Molecule 18; (B) scaffold 5, RMSD to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame.

Through π -sulphur interactions, the thioxozolidine ring's sulphur atom forms a 4.90 Å contact with the oxygen atom of tryptophan. As a result, the specificity pocket residue TRP111 (3.57 Å) causes π -sulphure and π - π stacking (3.80 Å) interactions with the thioxozolidine ring's sulphure atom. Likewise another specificity pocket residue PHE122 forms a π -suphur contact of 4.94 Å with sulphure atom of thioxozolidine ring of Molecule 18. TRP79 residues (5.53 Å) also create π -sulphure contacts with the sulphur atom of the thioxozolidine ring through π - π T shaped hydrophobic interactions with the naphthalene ring. Subsequently, specificity pocket residue LEU300 (5.03 Å) form π -alkyl hydrophobic contact with thioxozolidine ring, CYS303 anchored π -alkyl hydrophobic contact with thioxozolidine ring while VAL47 intercalate with naphthalene ring through π -alkyl hydrophobic interac-

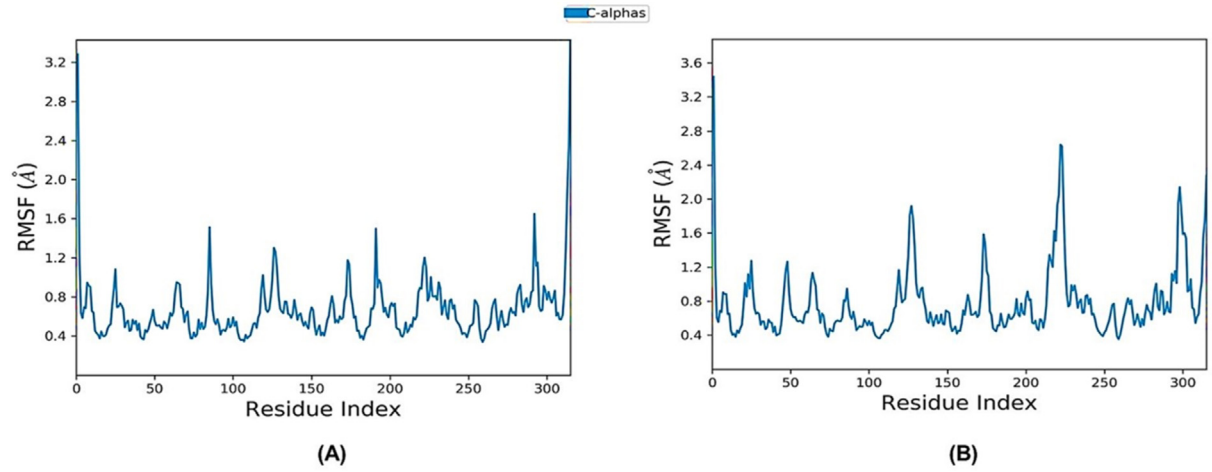


Fig. 19. Presentation of human AR – (A) Molecule 18; (B) scaffold 5 RMSF for characterizing local changes along the protein chain.

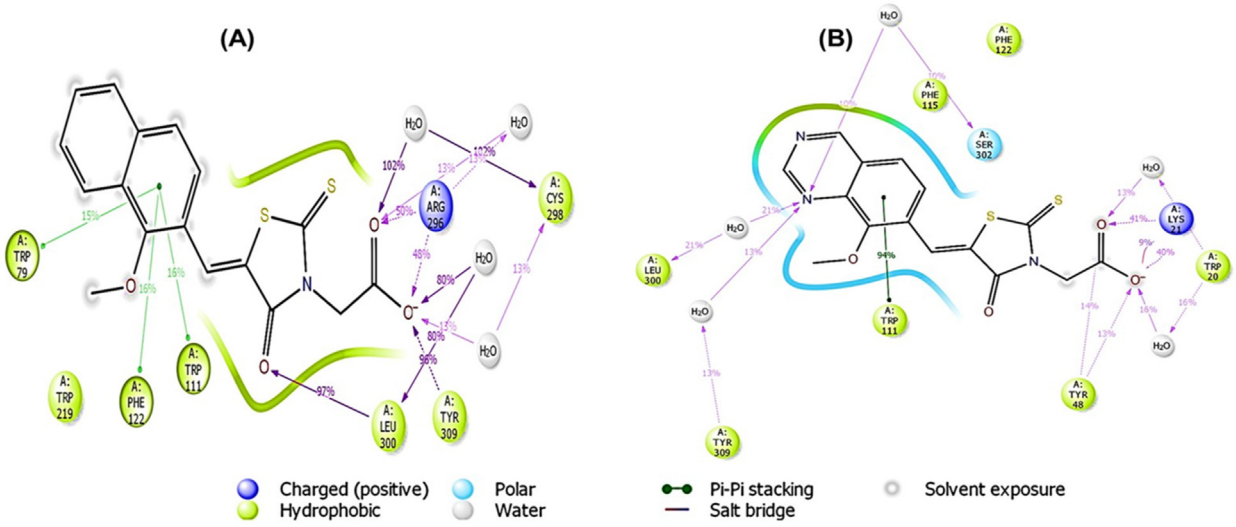


Fig. 20. 2D interaction plots showing ligand interactions of (A) Molecule 18; (B) Scaffold 5 with the binding cavity residues of Aldose Reductase (AR).

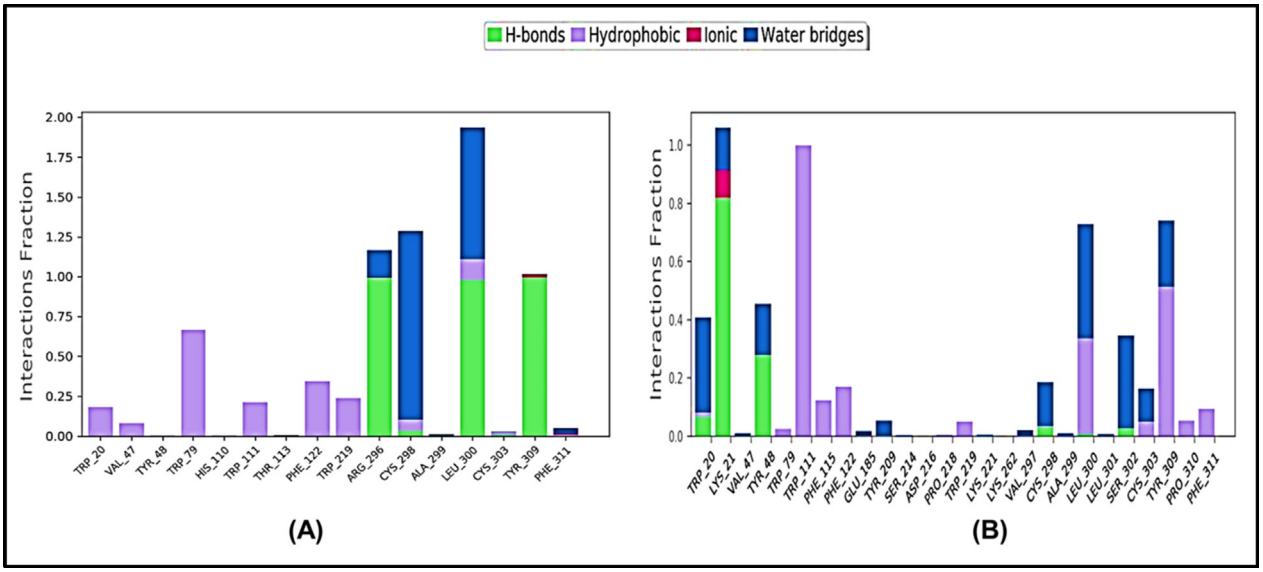


Fig. 21. (A) Molecule18 contact histogram (H-bonds, Hydrophobic, Ionic, Water bridges) of the ligand, molecule-18 bound with protein recorded in a 100 ns simulation interval; (B) Scaffold 5 contact histogram (H-bonds, Hydrophobic, Ionic, Water bridges) of the ligand, molecule-5 bound with protein recorded in a 100 ns simulation interval.

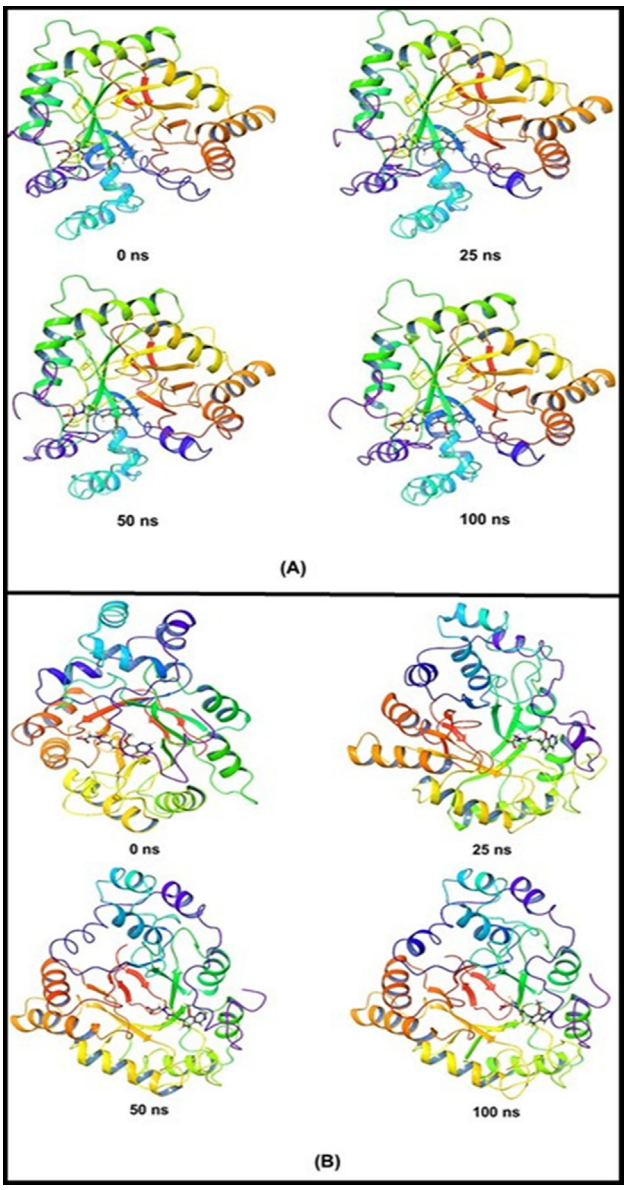


Fig. 22. Stepwise trajectory analysis for every 20 ns displaying the protein, (A) Scaffold 5 and molecule 18 and (B) Scaffold 5 and molecule 18; these show the conformation during 100 ns of simulation scale.

tions. Docking analysis revealed that Molecule 18 bind with specificity pocket residues Trp111, Ala299, Leu300 and Phe122 through conventional hydrogen bond, carbon hydrogen bond, π -sulphur interaction, π - π stacked hydrophobic and π -alkyl hydrophobic interactions. In Molecule 18, thioxozolidine ring, naphthalene ring, terminal carboxyl substituent, methoxy substituent on naphthalene ring plays crucial role for enhancing binding affinity against the human aldose reductase receptor.

4.11. Molecular docking analysis of scaffold 5.

As previously stated, the human AR receptor is divided into two sub-pockets, one containing residues that may be involved in catalysis (Tyr48, Lys77, and His110) that are sideways with the nicotinamide moiety of the cofactor, and the other containing Trp111, Ala299, Leu300, and Phe122 residues.

With a docking score of −8.08 kcal/mol, scaffold 5 achieved the same conformation in the specificity pocket as Pdb-1fzd ligand (see Fig. 16) and was anchored with specificity pocket residues Trp111, Ala299, and Leu300 via conventional hydrogen bonds, carbon hydrogen bonds, pi-sulfur contact, pi-donor hydrogen bond, π - π stacked hydrophobic contact, and π - π T-shaped contact, pi-alkyl contact and alkyl contacts.(see Fig. 17) According to the results of the molecular docking analysis, scaffold 5 has an excellent binding characteristic and affinity for the AR receptor. It also aligns into the specificity pocket in the same way the co-crystallized pdb ligand does.

The allminus SASA descriptor further emphasizes the need of a solvent-accessible surface. The presence of abundant SASA in the ligand molecule means that the ligand is exposed to the vanderwaals surface of the receptor molecule, increasing hydrophobic interactions. The network of solvent accessible surfaces of the ligand molecules around them remains intact as a result of hydrophobic groups tending to join together. This means that the results of QSAR and docking are complementary and consistent.

Furthermore, the Don_ringC_6Ac descriptor represents a blend of electronic and lipophilic features in the ligand molecule that contribute to hydrogen bonding and hydrophobic interactions. As a result, it is reasonable to conclude that the results of molecular docking and QSAR are complementary and in good agreement. As a result, the descriptor H_ringN_2B emphasizes the role of reduced steric bulk in enhancing receptor surface binding affinity. When we look at the docking score for ligand molecule 18 and scaffold 5, we can see that this observation is correct. As a result, the docking results match the QSAR analysis perfectly.

4.12. Molecular dynamics (MD) simulation study

MD simulation studies were performed in order to define the structural stability and conformational analysis of the screened nitrogen heterocycles with target protein is human aldose reductase (Pdb: 1fzd) which is a 36 kDa sized enzyme displays marked flexibility and malleability with respect to its active site. The Molecule 18 and scaffold 5 anchored with human aldose reductase and docking results were analyzed for 100 ns in MD simulation and displayed in Fig. 18. The conformation of Molecule 18 bound to the human aldose reductase displayed stable and converged conformation at the end of 100 ns (Fig. 18**.A**). Earlier till 50 ns a little RMSD fluctuations observed but later the system was fully converged with 1 Å deviation (Fig. 18 **A, blue**). On the other hand, ligand RMSD exhibited quite stable conformation from the beginning to end of the 100 ns simulation suggesting good fit and stable accommodation at the binding site of human aldose reductase protein (Fig. 18 **A, red**). The scaffold 5 bound to aldose reductase displayed stable conformation with less fluctuations in RMSD of both protein and ligand as displayed in Fig. 18.

The RMSF plots of Molecule 18 and scaffold 5 bound aldose reductase were displayed in Fig. 19. The C-α backbone displayed very less fluctuations in the respective amino acids positions with an average of 0.4 Å (Fig. 19 **(A)**). Whereas, scaffold 5 bound aldose reductase displayed significant fluctuations 1.2–2.0 Å between 110 and 120 residues and 2.8 Å between 210 and 220 residues (Fig. 19 **(B)**).

The interaction plots (Fig. 20 (A) and (B)) displayed the interaction of binding site residues of aldose reductase with the ligands Molecule 18 and scaffold 5. After 100 ns of simulation Molecule 18 formed conventional H-bonds with the Arg296, Leu300 and Tyr309 (Fig. 20 (A)) to entail into a stable complex. On the other hand, scaffold 5 binds by conventional H-bonds with Lys21 and pi-pi interaction with Trp111 (Fig. 20 (B)).

Throughout the simulation, protein interactions with the ligand can be observed. As seen in the graph above, these interactions can be classified and summarized by type. Hydrogen Bonds, Hydrophobic, Ionic, and Water Bridges are the four forms of protein–ligand interactions (or 'contacts'). Each interaction type has a number of

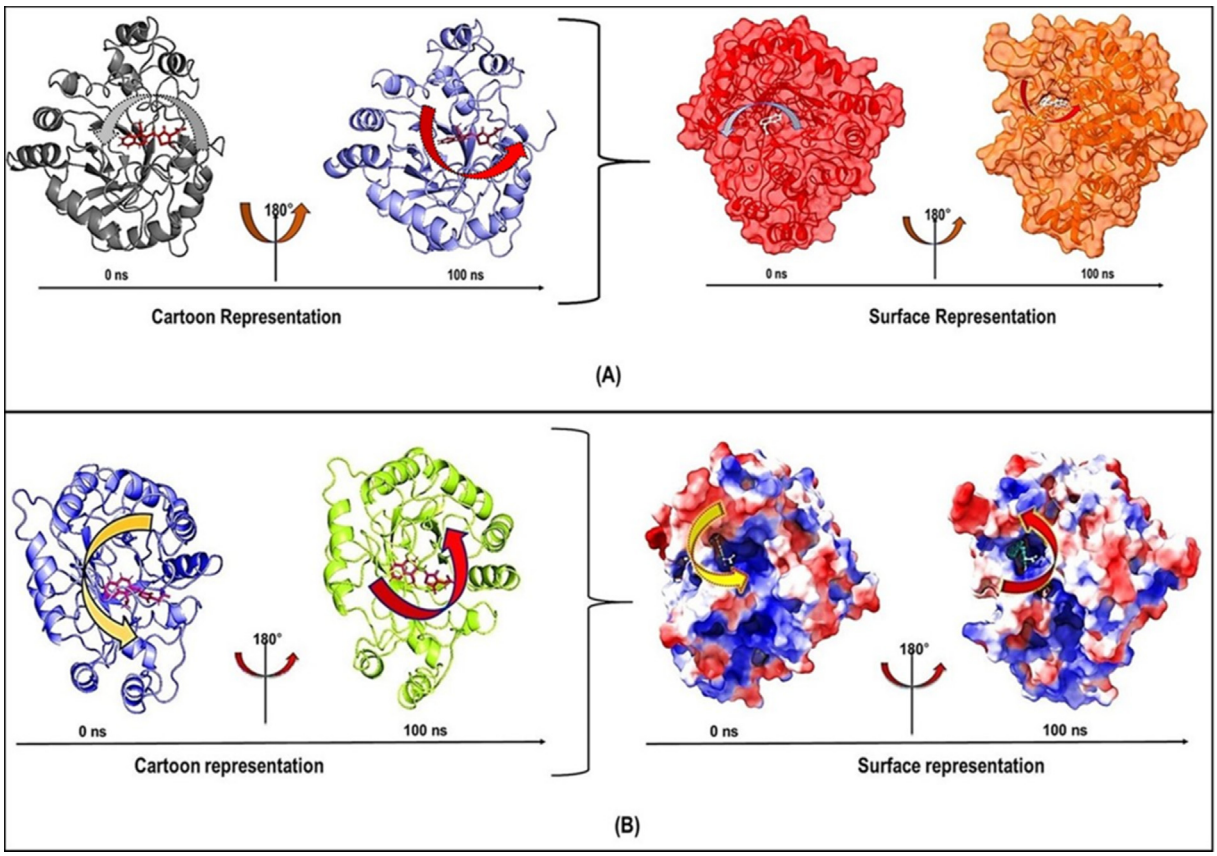


Fig. 23. MMGBSA trajectory (0 ns, before simulation and 100 ns, after simulation) exhibited conformational changes of molecule 18 and scaffold 5 upon binding with the protein, molecule 18 (A) and scaffold 5 (B). The arrows indicate the overall positional variation (movement and pose) of ligands at the binding site cavity.

Table 4
Binding energy calculation of molecule 18 and Scaffold 5 and non-bonded interaction energies from MMGBSA trajectories.

Energies (kcal/mol)	Molecule 18	Scaffold 5
ΔG_{bind}	−61.839 ± 5.673	−55.590 ± 3.705
ΔG_{bind}Lipo	−23.517 ± 1.693	−20.995 ± 1.509
ΔG_{bind}vdW	−49.459 ± 1.857	−43.737 ± 3.695
ΔG_{bind}Coulomb	17.834 ± 7.867	9.519 ± 3.212
$\Delta G_{bind}H_{bond}$	−1.829 ± 0.723	−1.128 ± 0.388
ΔG_{bind}SolvGB	16.579 ± 8.341	12.866 ± 10.172
ΔG_{bind}Covalent	1.362 ± 0.728	1.253 ± 0.801

subtypes that can be examined using Maestro's 'Simulation Interactions Diagram' panel, shown in Fig. 21. (A) and (B)). The stacked bar charts are standardized over the course of the trajectory. Some protein residues may make several interactions of the same subtype with the ligand, values above 1.0 are feasible. As seen in Fig. 21; the majority of the significant ligand–protein interactions discovered by MD are hydrogen bonds and hydrophobic interactions.

The stepwise trajectory analysis of every 25 ns of simulation of Scaffold 1 bound to molecule 18 and Scaffold 6 bound to molecule 5 displayed the positional alteration with reference to 0 ns structure [Fig. 22]. It has been observed that the ligand, molecule 18 with Scaffold 5 in Fig. 22 **(A)** and molecule 18 with Scaffold 5 in Fig. 22 **(B)** have possessed a structural angular movement at the end frame to achieve its conformational stability and convergence.

4.13. Molecular mechanics generalized born and surface area (MM GBSA) calculations

To assess the binding energy of ligands to protein molecules, the MMGBSA technique is commonly employed. The binding free energy of each molecule 18 complex and Scaffold 5, as well as the impact of other non-bonded interactions energies, were estimated. With Scaffold 1, the ligand molecule 18 has a binding energy of −61.836 kcal/mol, whereas with Scaffold 6, the ligand molecule 18 has a binding energy of −57.590 kcal/mol Non-bonded interactions like G_{bind}Coulomb, G_{bind}Covalent, G_{bind}Hbond, G_{bind}Lipo, G_{bind}SolvGB, and G_{bind}vdW govern Gbind. Across all types of interactions, the G_{bind}vdW, G_{bind}Lipo, and G_{bind}Coulomb energies contributed the most to the average binding energy. On the other side, the G_{bind}SolvGB and Gbind Covalent energies contributed the least to the final average binding energies. Furthermore, the G_{bind}Hbond interaction values of molecule 18 and Scaffold 5 complexes demonstrated stable hydrogen bonds with amino acid residues. In all of the compounds, G_{bind}SolvGB and G_{bind}Covalent exhibited unfavorable energy contributions, and so opposed binding. Fig. 23 **(A)** reveals that between pre-simulation (0 ns) and post-simulation (100 ns), molecule 18 in the binding pocket of Scaffold 1 has undergone a large angular change in the pose (curved to straight) and Fig. 23 **(B)** reveals that between pre-simulation (0 ns) and post-simulation (100 ns), molecule 5 in the binding pocket of Scaffold 6 has undergone a large angular change in the pose (curved to straight). These conformational changes lead to better binding pocket acquisition and interaction with residues, which leads to enhanced stability and binding energy (**shown in** Table 4).

Thus, MM-GBSA calculations resulted, from MD simulation trajectories well justified with the binding energy obtained from docking results moreover, the last frame (100 ns) of MMGBSA displayed the positional change of the molecule 18 and scaffold 5 as compared to 0 ns trajectory signify the better binding pose for best fitting in the binding cavity of the protein (**see** Fig. 23).

4.14. ADMET study of molecule 18 and scaffold 5

ADMET study of molecule 18 and scaffold 5 showed interesting molecular as well as pharmakokinetic properties. Total polar surface area of a molecule should be within the range of 20–130 Å. The molecule 18 TPSA was calculated to be 124.23 and for scaffold 5 TPSA is quite high 150 Å. The next descriptor (LogS, ESol model) was calculated to be water solubility whose range should be < 6. The molecule 18 showed high solubility having −4.5 whereas, scaffold 5 displayed −3.49. Therefore, it can be suggetsed that both molecule 18 and scaffold 5 have good solubility in water. GI absorption of molecule is found to be significant whereas, scaffold 5 has less GI absorption capacity. For drug likeliness property, molecule 18 displayed significant likeiness since none of the properties violate Lipinski, Ghose, Veber, Muegge and Egan filters. Whereas, scaffold 5, except Lipinski and Ghose all other filters are violated. Therefore, suggestively molceule 18 could be a proper drug where as scaffold has drug like properties. Molecule 18 has good capacity to tavel through BBB whereas, the possibility for scaffold 5 to pass through BBB is less. Moreover, both molecule 18 and scaffold 5 displayed to have good syntesizing accessibility score 3.37 and 3.2, respectively, which signifies both the lead molecules can be syntheiszed in the laboratory. As per the recommedation which should be < 6, both the molecule 18 and scaffold 5 are synthesizable. The overall comparative analysis of ADMET parmaters suggested that molecule 18 has better chance to become drug molecule instead scaffold 5.

5. Conclusions

In the present investigation, we used QSAR analysis, QSAR-based virtual screening, Molecular docking, and Molecular dynamic simulation to find a new AR receptor inhibitor that is a strong antidiabetic agent. To identify hidden structural features responsible for AR inhibition, a five parametric GA MLR-based full set and divided set QSAR model was developed. Pharmacophoric features such as solvent accessible surface area, partially charged ring carbon atoms within 6 Å of donor, combination of ring nitrogen and hydrogen separated by 2 bonds, and others emerged as projecting features that govern AR inhibition based on the developed QSAR model analysis. We conducted virtual screening using QSAR yielded a novel hit molecule (scaffold 5) with a pIC_{50} of 8.05 M (IC_{50} = 8.91 nM). Furthermore, molecular docking analysis of molecule 18 and scaffold 5 demonstrates that both molecules in the specificity binding pocket of AR adopted the same conformation as the pdb ligand. This study offered light on the pharmacophores involved in the binding interactions that inhibit AR. It gives the impression that, molecule18 and scaffold 5 bind with the specificity pocket residue A: LEU 300(1.90 Å, ALA299 (2.74 Å), ALA299 (2.62 Å), PRO310 (2.74 Å), TRP 20(5.21 Å). The QSAR and molecular docking results are consensus and complementary, and the identified pharmacophoric properties must be retained in the development of new and potential AR inhibitors as antidiabetic agents in the future. Finally, the high docking score alongside increase in IC_{50} value (8.91 nM) for scaffold 5 by 0.21 nm shows that it has a higher affinity for the specificity binding pocket of AR receptors, and this study could lead to the development of novel AR inhibitors as new antidiabetic agents. Subsequently, The formation of the significant number of hydrogen bonds in MD simulation corroborated the findings with molecular docking also suggested for a stable complex formation during the MD simulation over 100 ns time scale. MMGBSA is a powerful tool in determining the binding energy of the ligand with its respective protein targets. MMGBSA studies accurately predicted the total binding energy of theWedelosin at the binding cavity of ALK and BTK and exhibited a very low binding energy suggesting the capacity of theWe Scaffold 5 to conform into a stable complex. The binding energies in MMGBSA trajectory supported by van der Walls energy, Lipophilic energy, Coulombic energies and similarly reported elsewhere.The ADMET study revealed that scaffold 5 beheave as drug like candidate. The findings of this study may be relevant in the development of novel therapeutic targets for AR as an antidiabetic agent in the future.

Funding

The authors acknowledge the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University, Riyadh, KSA, for its support of this research through the Research group number 21–09-77. Authors are thankful to Dr. Paola Gramatica for providing free copy of QSARINS software.

CRediT authorship contribution statement

Ravindra L. Bakal: Conceptualization. **Rahul D. Jawarkar:** Conceptualization, Writing – original draft, Formal analysis. **J.V. Manwar:** Writing – original draft, Data curation. **M.S. Jaiswal:** Writing – original draft, Data curation. **Arabinda Ghosh:** . **Ajaykumar Gandhi:** . **Magdi E.A. Zaki:** Methodology, Writing – original draft. **Sami Al-Hussain:** . **Abdul Samad:** Writing – review & editing, Validation. **V.H. Masand:** Methodology, Writing – review & editing. **Nobendu Mukerjee:** . **Syed Nasir Abbas Bukhari:** . **Praveenkumar Sharma:** Writing – review & editing. **Israa Lewaa:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the Deanship of Scientific Research, Imam Mohammad Ibn Saud Islamic University, Saudi Arabia, Research Group no. RG-21-09-76. Authors are thankful to Dr. Paola Gramatica, for providing the free copy of QSARINS 2.2.4 Software. Authors are greatful to the R.C. Patel Institute of Pharmacy, Shirpur for providing computational facilities during entire course of research work. The Author, Rahul D. Jawarkar is thankful to The President, Shri Yogendraji Gode, for providing necessary facilities during Research work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsps.2022.04.003.

References

Bowers, K. J., D. E. Chow, H. Xu, et al., 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. ACM/IEEE SC 2006 Conference (SC'06): 43-43.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., 2014. QSAR Modeling: Where Have You Been? Where Are You Going To? J. Med. Chem. 57 (12), 4977–5010. https://doi.org/10.1021/jm4004285.

Consonni, V., R. Todeschini, D. Ballabio, et al., 2019. On the Misleading Use of QF32 for QSAR Model Comparison. Molecular Informatics. 38, Doi: 10.1002/minf.201800029.

Dearden, J.C., Cronin, M.T.D., Kaiser, K.L.E., 2009. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). SAR QSAR Environ. Res. 20 (3-4), 241–266. https://doi.org/10.1080/10629360902949567.
Foppiano, M., Lombardo, G., 1997. Worldwide pharmacovigilance systems and tolrestat withdrawal. The Lancet. 349 (9049), 399–400.
Fujita, T., Winkler, D.A., 2016. Understanding the Roles of the "Two QSARs". J. Chem. Inf. Model. 56 (2), 269–274. https://doi.org/10.1021/acs.jcim.5b00229.
Gramatica, P., 2013. On the Development and Validation of QSAR Models. Comput. Toxicol., 499–526
Gramatica, P., 2014. External Evaluation of QSAR Models, in Addition to Cross-Validation: Verification of Predictive Capability on Totally New Chemicals. Mol. Inf. 33 (4), 311–314. https://doi.org/10.1002/minf.201400030.
Gramatica, P., 2020. Principles of QSAR Modeling. Int. J. Quant. Struct. Property Relationships. 5 (3), 61–97. https://doi.org/10.4018/IJQSPR.20200701.oa1.
Gramatica, P., Cassani, S., Chirico, N., 2014. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. J. Comput. Chem. 35 (13), 1036–1044. https://doi.org/10.1002/jcc.23576.
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. J. Comput. Chem. 34 (24), 2121–2132. https://doi.org/10.1002/jcc.23361.
Harit, T., Bellaouchi, R., Asehraou, A., Rahal, M., Bouabdallah, I., Malek, F., 2017. Synthesis, characterization, antimicrobial activity and theoretical studies of new thiophene-based tripodal ligands. J. Mol. Struct. 1133, 74–79. https://doi.org/10.1016/j.molstruc.2016.11.051.
Jawarkar, R.D., Bakal, R.L., Zaki, M.E.A., Al-Hussain, S., Ghosh, A., Gandhi, A., Mukerjee, N., Samad, A., Masand, V.H., Lewaa, I., 2021. QSAR Based Virtual screening derived Identification of a Novel Hit as a SARS CoV-229E 3CLpro Inhibitor: GA-MLR QSAR modeling supported by Molecular Docking, Molecular Dynamics Simulation and MMGBSA calculation Approaches. Arabian J. Chem. 15 (1), 103499. https://doi.org/10.1016/j.arabjc.2021.103499.
Jez, J.M., Flynn, T.G., Penning, T.M., 1997. A new nomenclature for the aldo-keto reductase superfamily. Biochem. Pharmacol. 54 (6), 639–647.
Jorgensen, W.L., Maxwell, D.S., Tirado-Rives, J., 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc. 118 (45), 11225–11236. https://doi.org/10.1021/ja9621760.
Krans, H.M.J., 1993. Recent Clinical Experience With Aldose Reductase Inhibitors. Diabet. Med. 10, 44S–48S. https://doi.org/10.1111/j.1464-5491.1993.tb00198.x.
Kucerova-Chlupacova, M., Halakova, D., Majekova, M., Treml, J., Stefek, M., Soltesova Prnova, M., 2020. (4-Oxo-2-thioxothiazolidin-3-yl)acetic acids as potent and selective aldose reductase inhibitors. Chem. Biol. Interact. 332, 109286. https://doi.org/10.1016/j.cbi.2020.109286.
Martyna, G.J., Klein, M.L., Tuckerman, M., 1992. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. J. Chem. Phys. 97 (4), 2635–2643. https://doi.org/10.1063/1.463940.
Martyna, G.J., Tobias, D.J., Klein, M.L., 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101 (5), 4177–4189. https://doi.org/10.1063/1.467468.
Masand, V.H., Mahajan, D.T., Maldhure, A.K., Rastija, V., 2016. Quantitative structure–activity relationships (QSARs) and pharmacophore modeling for human African trypanosomiasis (HAT) activity of pyridyl benzamides and 3-(oxazolo[4,5-b]pyridin-2-yl)anilides. Med. Chem. Res. 25 (10), 2324–2334. https://doi.org/10.1007/s00044-016-1664-1.
Moon, H.-I., Jung, J.-C., Lee, J., 2006. Aldose reductase inhibitory effect by tectorigenin derivatives from Viola hondoensis. Bioorg. Med. Chem. 14 (22), 7592–7594. https://doi.org/10.1016/j.bmc.2006.07.002.
O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel: An open chemical toolbox. J. Cheminf. 3 (1). https://doi.org/10.1186/1758-2946-3-33.
Pavan, M., Todeschini, R., 2009. Multicriteria Decision-Making Methods. Comprehensive Chemometrics, 591–629.
Pourbasheer, E., Shokouhi Tabar, S., Masand, V.H., Aalizadeh, R., Ganjali, M.R., 2015. 3D-QSAR and docking studies on adenosine A2Areceptor antagonists by the CoMFA method. SAR QSAR Environ. Res. 26 (6), 461–477.
Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., Sherman, W., 2010. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. J. Chem. Theory Comput. 6 (5), 1509–1519. https://doi.org/10.1021/ct900587b.
Steele, J.W., Faulds, D., Goa, K.L., 1993. Epalrestat. Drugs Aging 3 (6), 532–555. https://doi.org/10.2165/00002512-199303060-00007.
Toukmaji, A.Y., Board, J.A., 1996. Ewald summation techniques in perspective: a survey. Comput. Phys. Commun. 95 (2-3), 73–92. https://doi.org/10.1016/0010-4655(96)00016-1.
Tsai, S.C., Burnakis, T.G., 2016. Aldose Reductase Inhibitors: An Update. Ann. Pharmacother. 27 (6), 751–754. https://doi.org/10.1177/106002809302700616.
van Gerven, J.M.A., Tjon-A-Tsien, A.M.L., 1995. The Efficacy of Aldose Reductase Inhibitors in the Management of Diabetic Complications. Drugs Aging 6 (1), 9–28. https://doi.org/10.2165/00002512-199506010-00002.
Bikadi, Z., Hazai, E., 2009. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. J. Cheminf. 1 (1). https://doi.org/10.1186/1758-2946-1-15.
Gaudreault, F., Morency, L.-P., Najmanovich, R.J., 2015. NRGsuite: a PyMOL plugin to perform docking simulations in real time using FlexAID. Bioinformatics. https://doi.org/10.1093/bioinformatics/btv458.
Gaulton, A., Hersey, A., Nowotka, M., Bento, A.P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L.J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M.P., Overington, J.P., Papadatos, G., Smit, I., Leach, A.R., 2017. The ChEMBL database in 2017. Nucleic Acids Res. 45 (D1), D945–D954. https://doi.org/10.1093/nar/gkw1074.
Ivanciuc, O., 1996. HyperChem Release 4.5 for Windows. J. Chem. Inf. Comput. Sci. 36 (3), 612–614. https://doi.org/10.1021/ci950190a.
Mukerjee, N., Das, A., Maitra, S., Ghosh, A., Khan, P., Alexiou, A., Dey, A., Baishya, D., Ahmad, F., Sachdeva, P., Al-Muhanna, M.K., Kumar, V., 2022. Dynamics of natural product Lupenone as a potential fusion inhibitor against the spike complex of novel Semliki Forest Virus. PLoS ONE 17 (2), e0263853. https://doi.org/10.1371/journal.pone.0263853.
Masand, V.H., Rastija, V., 2017. PyDescriptor : A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. Chemometr. Intell. Lab. Syst. 169, 12–18. https://doi.org/10.1016/j.chemolab.2017.08.003.