ACCSCIENCE PUBLISHING

Innovative Medicines & Omics

ORIGINAL RESEARCH ARTICLE

Integrated machine learning, molecular docking and ADMET analysis for computational prioritization of S1PR1-targeting drug candidates

Akshata P. Pauskar and Elamathi Natarajan*

Department of Bioinformatics, AI, and Machine Learning, Biotecnika Info Labs Pvt Ltd, Bangalore, Karnataka, India

Abstract

Sphingosine-1-phosphate receptor 1 (S1PR1) is a G protein-coupled receptor involved in immune-cell trafficking and has attracted considerable interest as a pharmacological target. The availability of structurally characterized ligand-bound S1PR1 provides an opportunity to combine ligand-based machine learning with structure-based screening to identify additional compounds with potential S1PR1 binding activity. This study aimed to develop and evaluate an integrated computational workflow to identify and prioritize compounds predicted to interact with S1PR1 from an external drug library. Specifically, we investigated whether molecular fingerprints and physicochemical descriptors could reliably predict S1PR1 activity and whether sequential machine-learning screening, drug-likeness filtering, molecular docking, interaction analysis, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction could prioritize candidates for further investigation. A curated ChEMBL dataset containing 3,300 unique compounds was used to develop random forest, extreme gradient boosting (XGBoost), logistic regression, and support vector machine models using scaffold-separated training and independent test sets. Random forest and XGBoost demonstrated strong predictive discrimination, with independent-test area under the receiver operating characteristic curve values of 0.9198 and 0.9252, respectively. The optimized random forest model was applied to 12,289 DrugBank compounds, of which 277 met the predefined prediction threshold. Subsequent Lipinski filtering retained 56 compounds for docking, with 54 yielding successful docking results. Integration of predicted activity, docking scores, physicochemical properties, protein–ligand interactions, and ADMET profiles prioritized five compounds: udifitimod, AZD8529, tedizolid phosphate, mocravimod, and BMS-986104. Chemical-space analysis indicated that many DrugBank compounds were structurally distant from the model-development chemical space. Overall, the workflow provides a computational strategy for prioritizing candidates for S1PR1-binding compounds for subsequent experimental evaluation.

Keywords: Sphingosine-1-phosphate receptor 1; Machine learning; Virtual screening; Molecular docking; ADMET; Drug discovery

***Corresponding author:**
Elamathi Natarajan
(elamathi@biotecnika.org)

Citation: Pauskar AP, Natarajan E. Integrated machine learning, molecular docking and ADMET analysis for computational prioritization of S1PR1-targeting drug candidates. *Innov Med Omics*. doi: 10.36922/IMO026300055

Received: July 20, 2026

Revised: October 7, 2026

Accepted: October 15, 2026

Published online: November 17, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The sphingosine-1-phosphate receptor 1 (S1PR1) is a master regulator of immune cell trafficking, controlling the movement of lymphocytes between lymphoid tissues and the bloodstream.[1] Its broad physiological roles in vascular integrity and inflammatory signaling mean its dysregulation has pleiotropic effects, contributing to a wide spectrum of autoimmune, cardiovascular, pulmonary, and oncological diseases.[2] In the specific context of multiple sclerosis (MS), this general function becomes a key disease mechanism.[3] S1PR1 enables the migration of autoreactive lymphocytes into the central nervous system (CNS), driving neuroinflammation, demyelination, and clinical progression. Consequently, pharmacological inhibition of S1PR1 has emerged as a breakthrough therapeutic strategy for MS, as it sequesters lymphocytes within lymph nodes, thereby reducing their infiltration into the CNS and mitigating damage (Figure 1).[4]

1.1. Mechanism of S1P–S1PR signaling pathway

Sphingosine-1-phosphate (S1P) is a bioactive sphingolipid metabolite that plays a central role in vascular homeostasis and immune cell trafficking. Intracellularly, it is generated from sphingosine by sphingosine kinase 1 within endothelial cells and other cell types.[5] Following its synthesis, S1P can either undergo irreversible degradation by S1P lyase in the endoplasmic reticulum or be exported into the extracellular space through specific transporters such as spinster homolog 2.[6] In the circulation, erythrocytes further contribute to maintaining high plasma S1P levels

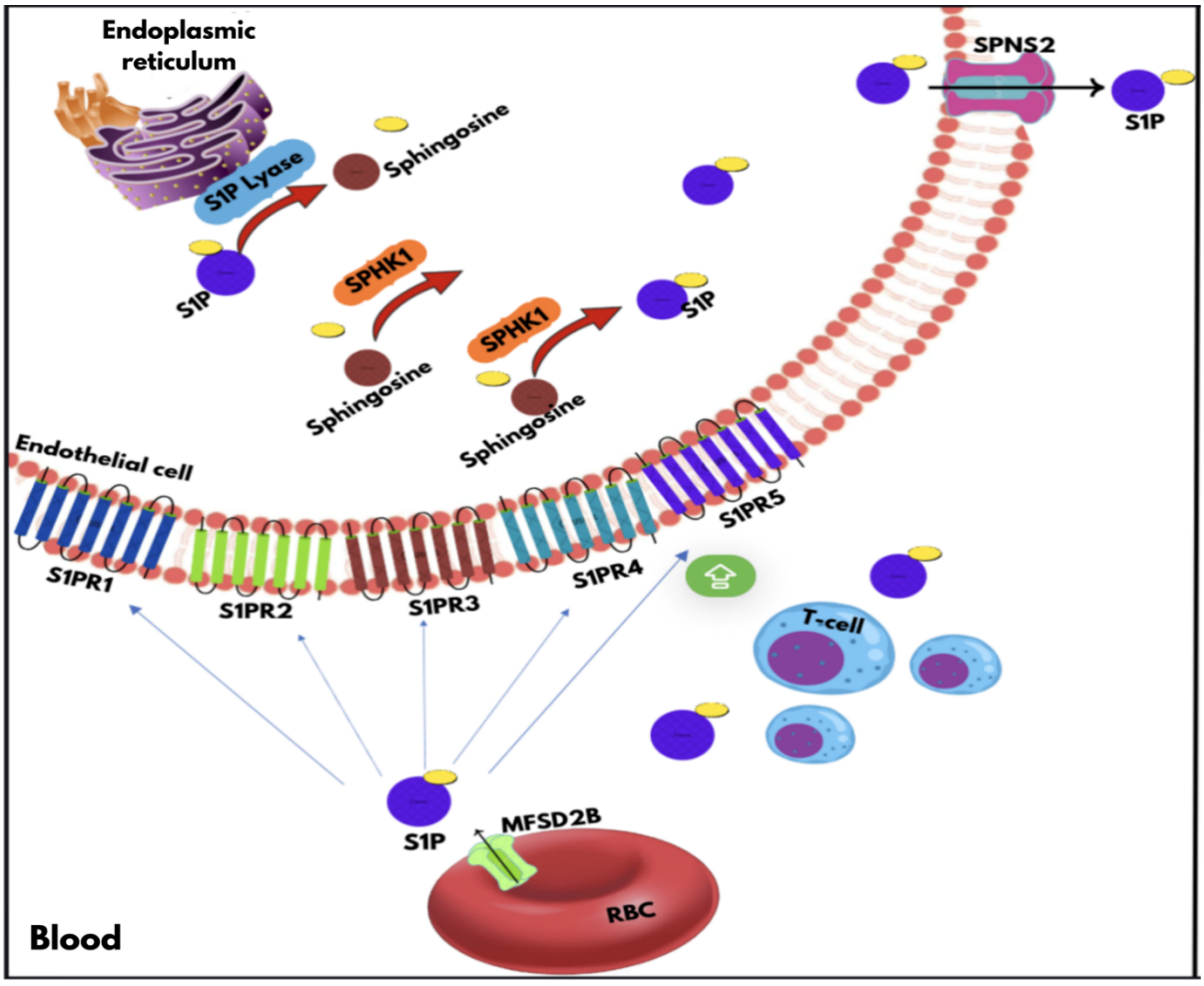


Figure 1. Schematic representation of sphingosine-1-phosphate (S1P) metabolism, transport, and signaling through S1P receptors (S1PR1–S1PR5) at the endothelial cell membrane.
Abbreviations: MFSD2B: Major facilitator superfamily domain-containing protein 2B; RBC: Red blood cell; S1P: Sphingosine-1-phosphate; S1PR: Sphingosine-1-phosphate receptor; SPHK1: Sphingosine kinase 1; SPNS: Spinster homolog.

via the major facilitator superfamily domain-containing protein 2B transporter, thereby establishing a steep S1P concentration gradient between blood and peripheral tissues. This gradient is fundamental for immune cell migration and vascular signaling.[7] Extracellular S1P exerts its biological effects by binding to a family of five G protein–coupled receptors (S1PR1–S1PR5), among which S1PR1 is particularly critical for endothelial integrity and lymphocyte trafficking.[8] Upon ligand binding, S1PR1 primarily couples to Gi proteins, activating downstream signaling pathways such as phosphoinositide 3-kinase/protein kinase B and Rac, which promote cytoskeletal rearrangement, adherens junction stabilization, and endothelial barrier enhancement.[9] In immune cells, especially T lymphocytes, S1PR1 senses the S1P gradient and regulates egress from lymphoid organs into the bloodstream.[10] Dysregulation of the S1P–S1PR1 axis has been implicated in inflammatory and autoimmune disorders, including MS, making it an important therapeutic target.[11] Modulation of S1PR1 signaling, as demonstrated by functional antagonists that induce receptor internalization and lymphocyte sequestration, highlights the clinical significance of this pathway in controlling immune-mediated pathology.[12]

1.2. Mechanism of S1PR1 modulators as functional antagonists in S1P–S1PR signaling

Clinically used S1PR1 modulators, including Fingolimod, Ozanimod, and Siponimod, operate by strategically hijacking the receptor's natural regulatory cycle.[13] Unlike the transient binding of the endogenous S1P ligand, these synthetic compounds act as functional antagonists by binding to S1PR1 with high affinity and triggering its pronounced and sustained internalization from the cell surface.[14] Once internalized, the drug-bound receptor is diverted from the normal recycling pathway and instead targeted for lysosomal degradation. This irreversible downregulation effectively depletes the lymphocyte's surface pool of S1PR1. Consequently, the cell is rendered "blind" to the physiological S1P gradient that normally guides its egress from lymphoid tissues. The result is the functional sequestration of lymphocytes, primarily autoreactive T and B cells, within lymph nodes, preventing their entry into systemic circulation and subsequent migration to sites of inflammation.[15] This mechanism creates a pharmacologically induced state of peripheral lymphopenia, which forms the basis for its therapeutic effect in autoimmune conditions (Figure 2).[16]

1.3. Accelerating drug discovery via artificial intelligence-enhanced screening

Traditional drug development is frequently hindered by high costs and high attrition rates.[17] To mitigate these risks, researchers utilize virtual screening (VS) to prioritize high-potential candidates for experimental validation.[18] Within this domain, structure-based VS is a cornerstone technique; it uses the 3D architecture of disease-associated proteins to model small-molecule interactions through molecular docking.[19,20]

The integration of artificial intelligence has further transformed this pipeline, enhanced predictive accuracy, and allowed for the rapid exploration of massive chemical libraries.[21] By refining docking simulations and combining them with molecular dynamics.[22,23] Artificial intelligence-driven workflows enable a more comprehensive understanding of binding stability and conformational changes, bridging the gap between static models and dynamic biological systems.[24]

We developed an integrated workflow that bridges machine learning with classical simulation techniques to discover potent S1PR1 inhibitors. By utilizing a trained ML model as an initial high-throughput filter, we efficiently narrowed a vast chemical space down to a refined set of top-ranked hits. These candidates were then validated through a secondary screening of docking and the evaluation of absorption, distribution, metabolism, excretion, and toxicity (ADMET). The sequential stages of this identification process are illustrated in the graphical abstract.

2. Methodology

2.1. Retrieval of protein structure

Sphingosine-1-phosphate receptor 1 was selected as the target protein for this study. The cryo-electron microscopy structure of human S1PR1 in complex with the agonist ozanimod (PDB ID: 7EW0) was obtained from the Protein Data Bank. This structure captures S1PR1 in its active conformation and was chosen as the reference model for structure-based docking studies, as it contains a bound agonist in the orthosteric binding pocket. The 7EW0 structure is organized as an asymmetric biological assembly with C1 symmetry and exhibits a hetero-pentameric stoichiometry (A1B1C1D1E1). It comprises five distinct protein chains, which include the S1PR1 receptor, subunits of the Gi protein, and additional associated components.

2.2. Identification of the active site

In the structural study reported by Yuan *et al.*[13], the cryo-electron microscopy structure of human S1PR1 in complex with Ozanimod (PDB ID: 7EW0) revealed the detailed architecture of the orthosteric binding pocket. The binding site was shown to comprise two major functional regions: (i) a polar headgroup interaction network located toward the extracellular side, formed by residues Y29, K34, R120[3.28],

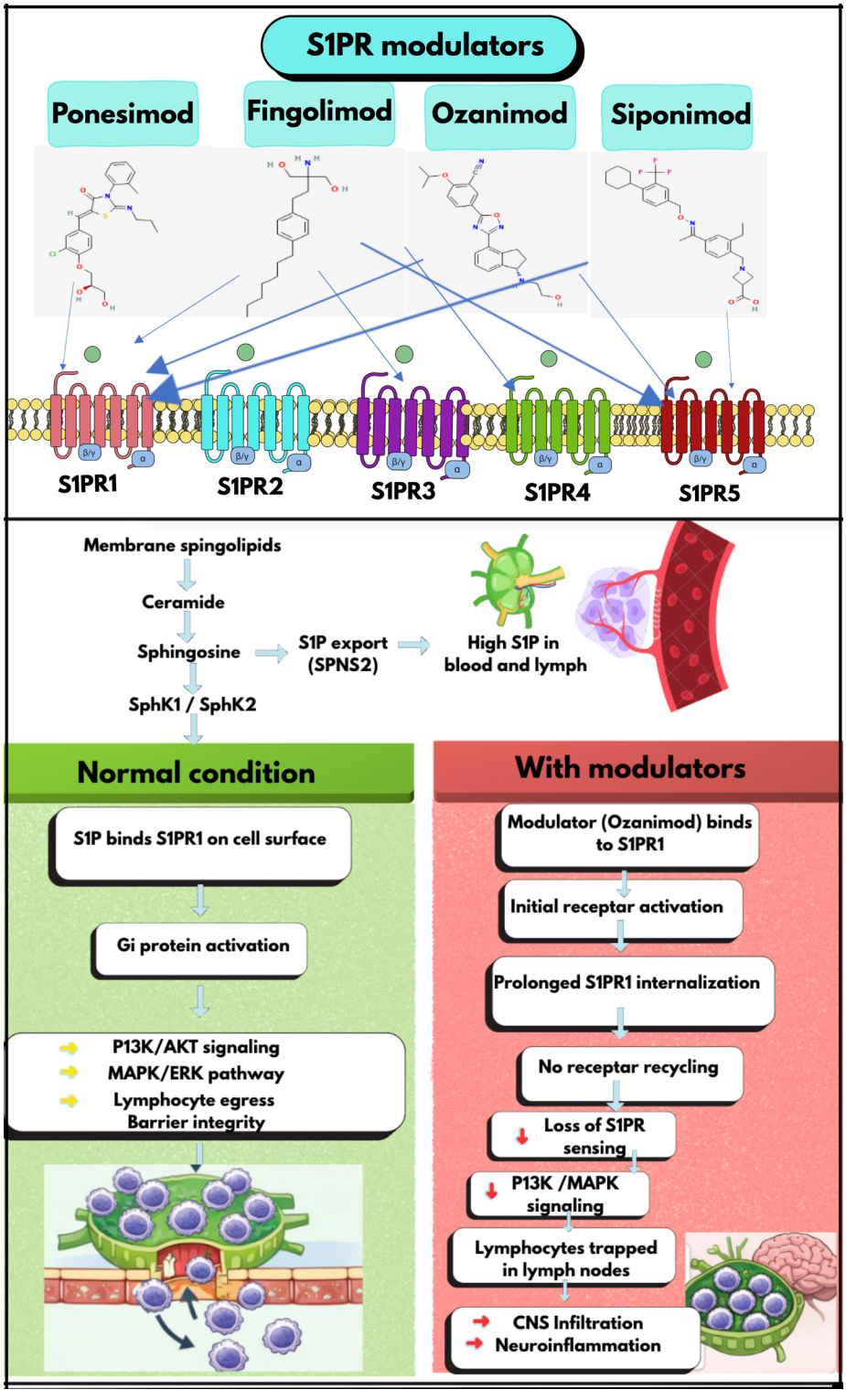


Figure 2. Schematic overview of S1P–S1PR1 signaling and the effect of S1PR modulators (ponesimod, fingolimod, siponimod, and ozanimod), illustrating S1PR1 internalization and reduced lymphocyte egress.

Abbreviations: AKT: Protein kinase B; CNS: Central nervous system; ERK: Extracellular signal-regulated kinase; MAPK: Mitogen-activated protein kinase; PI3K: Phosphoinositide 3-kinase; S1P: Sphingosine-1-phosphate; S1PR: Sphingosine-1-phosphate receptor; SPHK1: Sphingosine kinase 1; SPNS: Spinster homolog.

E121$^{3.29}$, and N101$^{2.60}$ (Ballesteros–Weinstein numbering for GPCR residues), which coordinate and stabilize the ligand's polar moiety through hydrogen bonding and electrostatic interactions; and (ii) a deep hydrophobic activation cavity extending within the transmembrane core, primarily constituted by L128$^{3.36}$, F210$^{5.47}$, W269$^{6.48}$, and F273$^{6.52}$, which accommodates the aromatic and hydrophobic portions of ozanimod. This dual interaction framework was demonstrated to be critical for ligand recognition, receptor activation, and subtype selectivity, thereby defining the structural basis of S1PR1 modulation described in their study. S1PR1 was selected as the target protein due to its critical role in immune regulation and its involvement in autoimmune diseases such as MS.[25]

2.3. Dataset preparation

Bioactivity data for S1PR1 (ChEMBL target ID: CHEMBL4333) were retrieved from the ChEMBL database. The initial dataset contained 11,625 bioactivity records corresponding to 4,566 unique ChEMBL compound identifiers. Exact duplicate records were removed prior to activity-based filtering. To obtain a consistent activity measure across compounds, records containing a non-missing standardized pChEMBL value were retained. This resulted in 4,426 records representing 3,423 unique compounds. The retained measurements comprised half-maximal effective concentration (EC50) (n = 2,909), half-maximal inhibitory concentration (IC50) (n = 1,372), inhibition constant (n = 144), and dissociation constant (n = 1) measurements. Because the present study focused on functional S1PR1 activity, only EC50 and IC50 measurements were retained for subsequent machine-learning model development. This filtering resulted in 4,281 records corresponding to 3,324 unique compounds.

Activity classes were assigned using the standardized pChEMBL value.[26,27] Compounds with pChEMBL ≥ 6 were classified as active, whereas compounds with pChEMBL < 6 were classified as inactive. A pChEMBL value of 6 corresponds approximately to an activity concentration of 1 μM and was therefore used as an operational threshold to distinguish compounds exhibiting activity at or below the micromolar range from compounds with weaker measured activity. The classification criterion was applied consistently to the retained EC50 and IC50 measurements. Compounds for which measurements produced conflicting activity classes were excluded rather than assigning an arbitrary class. Among the 3,324 compounds represented by the retained functional measurements, 24 compounds showed conflicting active/inactive classifications and were removed.

The resulting curated dataset contained 4,228 measurements corresponding to 3,300 unique compounds, comprising 2,718 active compounds and 582 inactive compounds. For compounds represented by multiple retained measurements, the median pChEMBL value was retained as the compound-level activity value. The final curated dataset contained no missing or empty Simplified Molecular Input Line Entry System (SMILES) strings and no shared structures between different ChEMBL compound identifiers based on canonical SMILES comparison.

2.4. Molecular structure processing and data splitting

Bemis–Murcko scaffolds were generated for all curated compounds using RDKit (2026.3.6, Rational Discovery, USA). To minimize structural information leakage between model-development and evaluation datasets, compounds were divided using a scaffold-based 80:20 split. The resulting training set contained 2,399 compounds comprising 2,015 active and 384 inactive compounds, and the independent test set contained 901 compounds comprising 703 active and 198 inactive compounds. No Bemis–Murcko scaffolds or InChIKeys were shared between the training and test sets, providing a scaffold-independent evaluation of model performance. Activity labels were encoded as 0 for inactive compounds and 1 for active compounds. The final dataset was used for subsequent feature generation and machine learning model development.[28–30]

2.5. Molecular feature generation

Molecular structures were represented using circular Morgan fingerprints and physicochemical descriptors. Morgan fingerprints equivalent to Extended-Connectivity Fingerprint with a diameter of 4 (ECFP4) were generated using a radius of 2 and a fingerprint length of 1,024 bits using RDKit's Morgan fingerprint generator. Eleven molecular descriptors were additionally calculated: molecular weight (MW), calculated lipophilicity (logP), hydrogen-bond acceptor (HBA) count, hydrogen-bond donor (HBD) count, rotatable bond count, heavy-atom count, topological polar surface area, aromatic ring count, fraction of sp3-hybridized carbons, total ring count, and number of valence electrons. The resulting feature representation consisted of 1,035 molecular features per compound (1,024 fingerprint features plus 11 molecular descriptors).[31–33]

2.6. Machine-learning model development and hyperparameter optimization

Four supervised machine-learning algorithms were evaluated: Random forest (RF), support vector machine (SVM) with a radial basis function (RBF) kernel, logistic regression (LR), and extreme gradient boosting (XGBoost).

To address class imbalance while preventing information leakage, SMOTE was incorporated within an imbalanced-learn pipeline and was applied only to the training portion of each cross-validation fold. The independent test set was never subjected to SMOTE. Standardization was additionally applied within the SVM-RBF and logistic-regression pipelines.[34–36]

Hyperparameters were optimized using RandomizedSearchCV with 20 randomly sampled parameter combinations for each model. Five-fold stratified cross-validation was performed with shuffling and a fixed random seed. The optimized models were subsequently refit using the complete training dataset and evaluated on the independent scaffold-separated test set.

2.7. Model performance evaluation

Model performance was assessed using receiver operating characteristic (ROC)-area under the curve (AUC), precision–recall AUC (PR-AUC), accuracy, sensitivity, specificity, precision, F1-score, balanced accuracy, and Matthews correlation coefficient (MCC). Confusion matrices were generated to examine the numbers of true positives, true negatives, false positives, and false negatives.[37] The mean ROC-AUC and standard deviation across the five cross-validation folds were calculated. Approximate 95% confidence intervals (CIs) were calculated from the five-fold-level ROC-AUC values using the corresponding t-distribution. Because VS requires prioritization of compounds according to predicted activity, RF and XGBoost probability thresholds ranging from 0.10 to 0.95 were additionally examined. Sensitivity, specificity, precision, F1-score, balanced accuracy, MCC, and the number of compounds classified as active were calculated at each threshold.

Probability calibration was assessed on the independent test set using calibration curves and the Brier score. Lower Brier scores indicate smaller differences between predicted probabilities and observed binary outcomes, and the trained model was subsequently used for VS.[38]

2.8. DrugBank as screening dataset

A second dataset comprising 12,309 compounds was retrieved from DrugBank in SDF format for prospective VS. The compound name and molecular structure (SMILES) were retained for subsequent analysis. Molecular structures were processed using RDKit. Compounds with missing or empty SMILES representations were removed, and duplicate structures were identified using canonical SMILES. This procedure resulted in 12,300 unique molecular structures. All retained structures were subsequently subjected to RDKit molecular validation, and no invalid molecular structures were detected.

To prevent exact chemical overlap between the external screening library and the compounds used for model development and evaluation, InChIKeys were generated for the DrugBank compounds and compared with the InChIKeys of the ChEMBL training and independent test sets. Six DrugBank compounds showed exact structural identity with compounds in the training set, while five showed exact identity with compounds in the independent test set. These 11 overlapping compounds were excluded, resulting in a final DrugBank screening library of 12,289 compounds. This exact-structure filtering was performed before machine-learning prediction to minimize potential information leakage between the model-development dataset and the prospective screening library.

2.9. Molecular feature generation for DrugBank screening

The same molecular representation used during model development was applied to the DrugBank screening library to ensure feature consistency between the training and external screening datasets. For each DrugBank compound, a 1,024-bit Morgan fingerprint with radius 2 (ECFP4-equivalent representation) was generated using RDKit.[39] In addition, the same 11 physicochemical and structural descriptors used for the training set were calculated. The resulting fingerprint and descriptor matrices were concatenated to produce a 1,035-dimensional molecular feature representation for each DrugBank compound, matching the feature representation used for the ChEMBL training and test datasets.

2.10. Machine-learning-based DrugBank virtual screening

The optimized machine-learning models generated during model development were applied to the 12,289-compound DrugBank library. RF, XGBoost, LR, and SVM models were evaluated using their respective optimized pipelines. For prospective screening, the RF model was selected as the operational screening model. The model incorporated the same SMOTE-based training pipeline used during model development.[40]

The RF model generated a continuous predicted activity score for each DrugBank compound. A score of ≥ 0.70 was used as the operational threshold for selecting compounds for subsequent structure-based filtering and molecular docking. This threshold was selected based on the independent ChEMBL test-set threshold analysis, where increasing the threshold from 0.50 to 0.70 improved specificity from 0.717 to 0.853 and precision from 0.922 to 0.954, while retaining a sensitivity of 0.859. The threshold

was used to prioritize compounds for downstream computational analysis.

2.11. Lipinski rule of five

The compounds identified from the RF VS step were further evaluated for drug-like physicochemical properties using Lipinski's Rule of Five.[41] The filtering criteria applied were MW ≤ 500 Da, calculated logP ≤ 5, HBD count ≤ 5, and HBA count ≤ 10. Compounds satisfying all four criteria were retained for subsequent molecular preparation and docking analysis. In this study, the Lipinski criteria were applied as a drug-likeness prioritization filter and were not considered evidence of biological activity or favorable pharmacokinetics.[42,43]

2.12. Molecular docking and redocking validation

The Lipinski-filtered compounds were converted from canonical SMILES into RDKit molecular objects. Explicit hydrogen atoms were added to each molecular structure, followed by generation of three-dimensional coordinates using the RDKit EmbedMolecule function with a fixed random seed. The generated molecular geometries were subsequently subjected to Universal Force Field geometry optimization using RDKit. The generated ligand SDF files were converted into PDBQT format using Open Babel (version 3.1.1, OpenEye Scientific Software, USA).[44]

Molecular docking was performed using AutoDock Vina (version 1.2.5, The Scripps Research Institute, USA) to evaluate the binding poses and predicted docking scores of the prioritized compounds against the S1PR1 receptor.[45] The crystal structure of S1PR1 complexed with JEU (PDB ID: 7EW0) was used as the receptor structure.[13] Water molecules and the crystallographic ligand were removed, and the receptor structure was separated and prepared using the DockPrep tool in UCSF Chimera (version 1.19, University of California, USA). The prepared receptor was converted to PDBQT format for docking.[46,47] The co-crystallized ligand's (JEU) spatial coordinates were extracted using the "*get_extent*" function in PyMOL (version 3.1.6.1, Schrödinger, Inc., USA).

Prior to screening the selected compounds, the docking protocol was assessed by redocking the crystallographic JEU ligand into the prepared S1PR1 binding site. The docking grid was centered at 107.075, 112.817, and 74.658 Å along the X, Y, and Z axes, respectively, with dimensions of 17.5 × 7.0 × 19.0 Å. JEU redocking was performed using AutoDock Vina with exhaustiveness values of 8, 16, and 32, and a fixed random seed of 123456.[48,49] The top-ranked redocked pose was compared with the crystallographic JEU conformation using heavy-atom root mean square deviation (RMSD).

Based on the redocking assessment, an exhaustiveness value of 16 was used for subsequent docking because the results obtained at exhaustiveness 16 and 32 were identical in both the best docking score and top-pose RMSD. The compounds that passed the machine-learning and Lipinski filtering stages were docked against the same prepared receptor and within the same grid box. Docking was performed with an exhaustiveness of 16 and a fixed random seed of 123456. The best-ranked Vina score for each successfully docked compound was recorded for subsequent prioritization. Compounds with predicted docking scores of ≤ −7.5 kcal/mol were retained. This threshold was used only for ranking and was not interpreted as an experimentally validated binding-affinity cutoff. Following docking, the top-ranked pose from each docking output was extracted and converted from PDBQT to PDB format using Open Babel for subsequent visualization and interaction analysis.

2.13. Protein–ligand interaction analysis

To further evaluate the docking poses of the computationally prioritized compounds, protein–ligand interaction analysis was performed using the S1PR1 structure PDB ID 7EW0, which contains ozanimod (JEU) as the crystallographic reference ligand. The crystallographic ozanimod binding environment was first examined to define the experimentally observed ligand-binding region. The analysis encompassed identification and classification of hydrogen bonds, determination of key amino acid residues participating in ligand binding within the receptor pocket, and measurement of intermolecular bond distances.[50]

The interaction profiles of the five compounds were compared with those of the crystallographic ozanimod ligand and its redocked pose. The redocking of ozanimod produced an RMSD of 1.049 Å, providing a reference for the predicted binding orientation within the 7EW0 binding site. The interaction analysis was used as a complementary structural criterion alongside ML prediction probability, docking score, physicochemical properties, and ADMET predictions for computational prioritization.

2.14. Drug-likeness and ADMET analysis

Following molecular docking, the five computationally prioritized S1PR1 candidates were further evaluated for physicochemical properties and drug-likeness. MW, logP, number of rotatable bonds, HBA, HBD, and topological surface area were calculated using the same computational workflow used for candidate characterization. These parameters were assessed with reference to Lipinski's Rule of Five, considering MW ≤ 500 Da, logP ≤ 5, HBD ≤ 5, and HBA ≤ 10 as the principal criteria for oral drug-likeness.

The crystallographic reference ligand, ozanimod (JEU), was included for comparison.

The ADMET properties of the selected compounds were predicted using the pkCSM web server. Predicted absorption, distribution, metabolism, excretion, and toxicity parameters were evaluated to identify potential pharmacokinetic characteristics and toxicity-related liabilities.[51,52] The analysis included water solubility, Caco-2 permeability, human intestinal absorption, P-glycoprotein substrate status, volume of distribution, fraction unbound, blood–brain barrier/CNS permeability, selected cytochrome P450 (CYP)-mediated metabolic interactions, total clearance, renal organic cation transporter 2 substrate status, Ames toxicity, human ether-à-go-go-related gene inhibition, and hepatotoxicity. The complete set of predicted pkCSM parameters was retained as supplementary information.

2.15. Chemical-space similarity and applicability-domain assessment

To assess the chemical-space relationship between the external DrugBank screening library and the compounds used for machine-learning model development, molecular fingerprint-based similarity analysis was performed using RDKit. Morgan fingerprints equivalent to ECFP4 were generated using a radius of 2 and a fingerprint length of 1,024 bits, consistent with the molecular representation used for model development.[53,54]

For each compound in the DrugBank screening library, its Morgan fingerprint was compared with the fingerprints of all 2,399 compounds in the ChEMBL training set using the Tanimoto similarity coefficient. The maximum Tanimoto similarity to any training compound was retained for each DrugBank molecule as a measure of its closest structural relationship to the model-development chemical space. Summary statistics and the proportions of compounds exceeding Tanimoto similarity thresholds of 0.3, 0.4, 0.5, 0.6, and 0.7 were calculated.

As an internal reference for interpreting the chemical-space relationship, the same procedure was applied to the 901 compounds in the independent ChEMBL test set, with each test compound compared against all 2,399 training compounds. The resulting maximum-similarity distributions were compared with those obtained for the DrugBank screening library.

3. Results and discussion

3.1. Machine-learning model performance

The curated dataset comprised 3,300 unique compounds, including 2,718 active and 582 inactive compounds. The scaffold-separated training and independent test sets contained 2,399 and 901 compounds, respectively. Across five-fold cross-validation, all four models showed good discriminatory performance (Table 1). RF achieved a mean ROC-AUC of 0.9230 ± 0.0184 (95% CI: 0.9001–0.9458), followed by XGBoost (0.9182 ± 0.0220, 95% CI: 0.8909–0.9456), LR (0.9031 ± 0.0159, 95% CI: 0.8833–0.9229), and SVM-RBF (0.8866 ± 0.0270, 95% CI: 0.8530–0.9202).

3.2. Independent test-set evaluation

Evaluation on the independent scaffold-separated test set demonstrated that all four models retained good discriminative ability when evaluated on compounds not represented by the corresponding training scaffolds. The ROC and precision–recall analyses showed consistent separation between active and inactive compounds across the models, although their performance characteristics differed.

The RF and XGBoost showed better overall performance, with high accuracy and ROC-AUC values and consistently high precision, recall, and F1-scores for the active class (Table 2). RF showed a slightly higher accuracy, whereas XGBoost achieved the highest ROC-AUC. LR showed comparatively stronger recall for the inactive class but lower precision. SVM-RBF showed the lowest overall performance among the evaluated models, particularly in classifying the inactive class.

The confusion matrices (Figure 3) showed that RF correctly classified 662 active and 142 inactive compounds, with 41 false negatives and 56 false positives. XGBoost correctly classified 653 active and 141 inactive compounds, with 50 false negatives and 57 false positives. LR produced

Table 1. Five-fold cross-validation performance of the machine learning models

Model	Random forest	XGBoost	Logistic regression	SVM-RBF
Mean ROC-AUC	0.9230	0.9182	0.9031	0.8866
SD	0.0184	0.0220	0.0159	0.0270
95% CI	0.9001–0.9458	0.8909–0.9456	0.8833–0.9229	0.8530–0.9202

Abbreviations: AUC: Area under the curve; CI: Confidence interval; RBF: Radial basis function; ROC: Receiver operating characteristic; SD: Standard deviation; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

fewer false positives (29) but a greater number of false negatives (113), while SVM-RBF produced 78 false positives and 66 false negatives.

Random forest and XGBoost showed the best overall discrimination, with closely comparable performance. XGBoost provided the highest ranking-based discrimination, whereas RF showed a better balance between sensitivity and specificity at the evaluated classification threshold (Table 3). In particular, the higher sensitivity of RF indicates that it identified a larger proportion of active compounds while maintaining moderate specificity, which is relevant for a screening-oriented application where reducing the number of missed active candidates is important.

The LR displayed a different classification profile, with stronger discrimination of inactive compounds but comparatively lower sensitivity. SVM-RBF showed the weakest overall discrimination among the evaluated models and a lower ability to exclude inactive compounds at the selected threshold.

Overall, the independent test-set results demonstrated that RF and XGBoost maintained good predictive discrimination on chemically distinct scaffolds, supporting their subsequent application to VS. The differences in sensitivity, specificity, and MCC also indicated that model performance depended on the classification threshold and could not be characterized adequately using ROC-AUC alone. Therefore, probability-threshold analysis was subsequently performed for RF and XGBoost to examine their suitability for compound prioritization.

The ROC curves generated using the independent test set demonstrated discriminatory performance above the random-classifier baseline for all four models (Figure 4). RF achieved a test-set ROC-AUC of 0.9198, while

Table 2. Overall performance comparison table of machine learning models on the independent test set for the inactive (Class 0) and active (Class 1) compounds

Model		Random forest	XGBoost	Logistic regression	SVM-RBF
Accuracy		0.8923	0.8812	0.8424	0.8402
ROC-AUC		0.9198	0.9252	0.9032	0.8667
Class 0 (inactive)	Precision	0.7760	0.7382	0.5993	0.6452
	Recall	0.7172	0.7121	0.8535	0.6061
	F1 score	0.7454	0.7249	0.7042	0.6250
Class 1 (active)	Precision	0.9220	0.9197	0.9532	0.8909
	Recall	0.9417	0.9289	0.8393	0.9061
	F1 score	0.9317	0.9243	0.8926	0.8984

Abbreviations: AUC: Area under the curve; RBF: Radial basis function; ROC: Receiver operating characteristic; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

Table 3. Independent test-set performance of the machine learning models

Model	Random forest	XGBoost	Logistic regression	SVM-RBF
ROC-AUC	0.9198	0.9252	0.9032	0.8667
PR-AUC	0.9757	0.9771	0.9675	0.9569
Accuracy	0.8923	0.8812	0.8424	0.8402
Sensitivity	0.9417	0.9289	0.8393	0.9061
Specificity	0.7172	0.7121	0.8535	0.6061
Precision	0.9220	0.9197	0.9532	0.8909
F1	0.9317	0.9243	0.8926	0.8984
Balanced accuracy	0.8294	0.8205	0.8464	0.7561
MCC	0.6781	0.6494	0.6187	0.5240
Brier score	0.0917	0.0920	0.1239	0.1292

Abbreviations: AUC: Area under the curve; MCC: Matthews correlation coefficient; PR: Precision–recall; RBF: Radial basis function; ROC: Receiver operating characteristic; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

XGBoost produced the highest independent-test ROC-AUC of 0.9252. LR and SVM-RBF produced ROC-AUC values of 0.9032 and 0.8667, respectively.

The close performance of RF and XGBoost indicates that both models yielded better classification results on the structurally independent test set, although their relative performance differed slightly between cross-validation and independent-test evaluation.

3.3. Probability calibration

Calibration analysis on the independent test set gave Brier scores of 0.0917 for RF, 0.0920 for XGBoost, 0.1239 for LR, and 0.1292 for SVM-RBF (Figure 4). RF and XGBoost therefore showed similar probability-calibration performance, whereas higher Brier scores were observed for LR and SVM-RBF.

The calibration curves showed deviations from the perfect calibration diagonal, particularly within intermediate probability ranges. Therefore, predicted probabilities were treated primarily as model-derived screening scores rather than as directly calibrated probabilities of biological activity.

3.4. Threshold analysis for virtual screening

Because the ultimate application of the model was VS rather than binary classification alone, probability thresholds were examined using the independent test set.

For RF, increasing the threshold from 0.50 to 0.70 increased specificity from 0.717 to 0.853 and precision from 0.922 to 0.954, while sensitivity decreased from 0.941 to 0.859. At a threshold of 0.70, the RF model achieved a balanced accuracy of 0.856 with an MCC of 0.645 and classified 633 of the 901 test compounds as active. At the same threshold, XGBoost showed a sensitivity of 0.920, specificity of 0.782, precision of 0.937, balanced accuracy of 0.851, and MCC of 0.687. The 0.70 RF threshold was

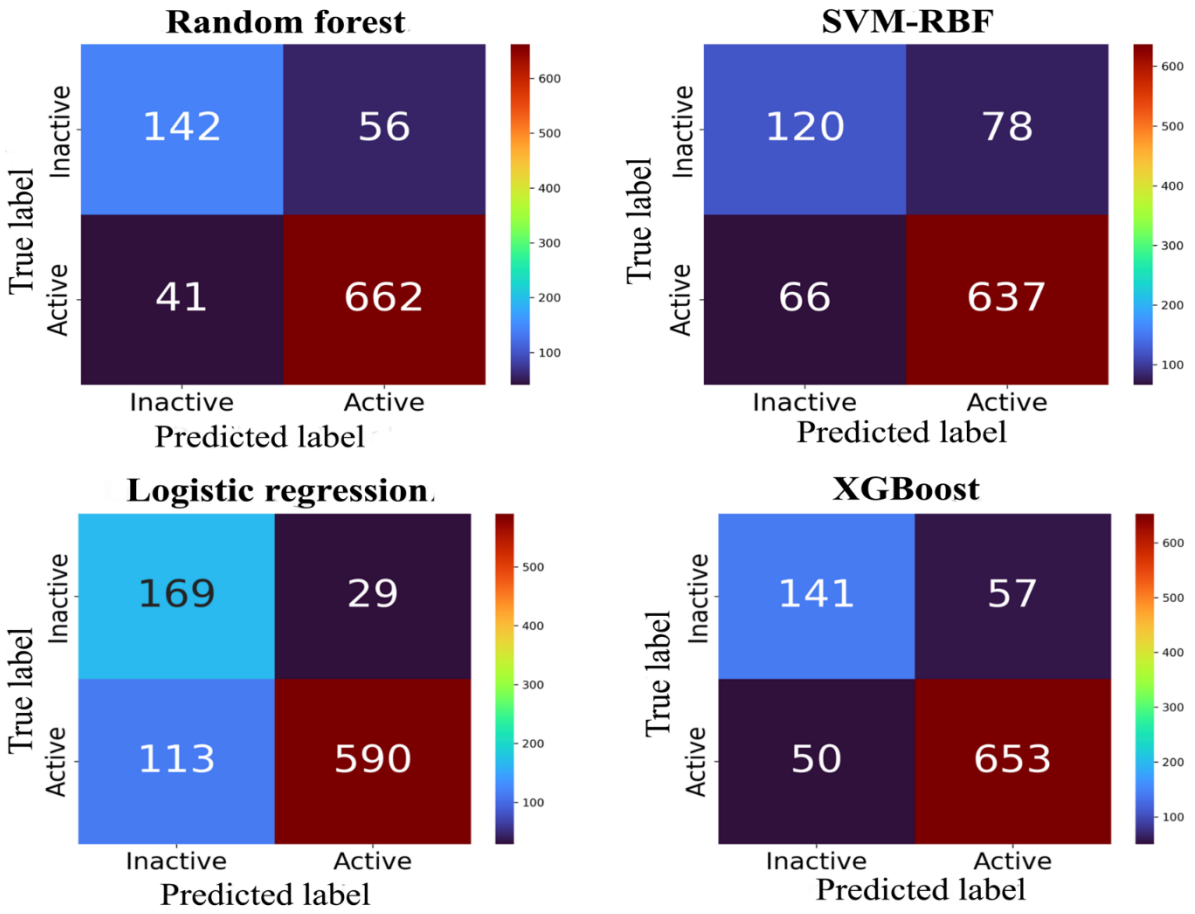


Figure 3. Confusion matrices of the machine learning models on the independent test set. Confusion matrices showing the classification performance of random forest, support vector machine (SVM) with radial basis function (RBF) kernel, logistic regression (LR), and extreme gradient boosting (XGBoost).

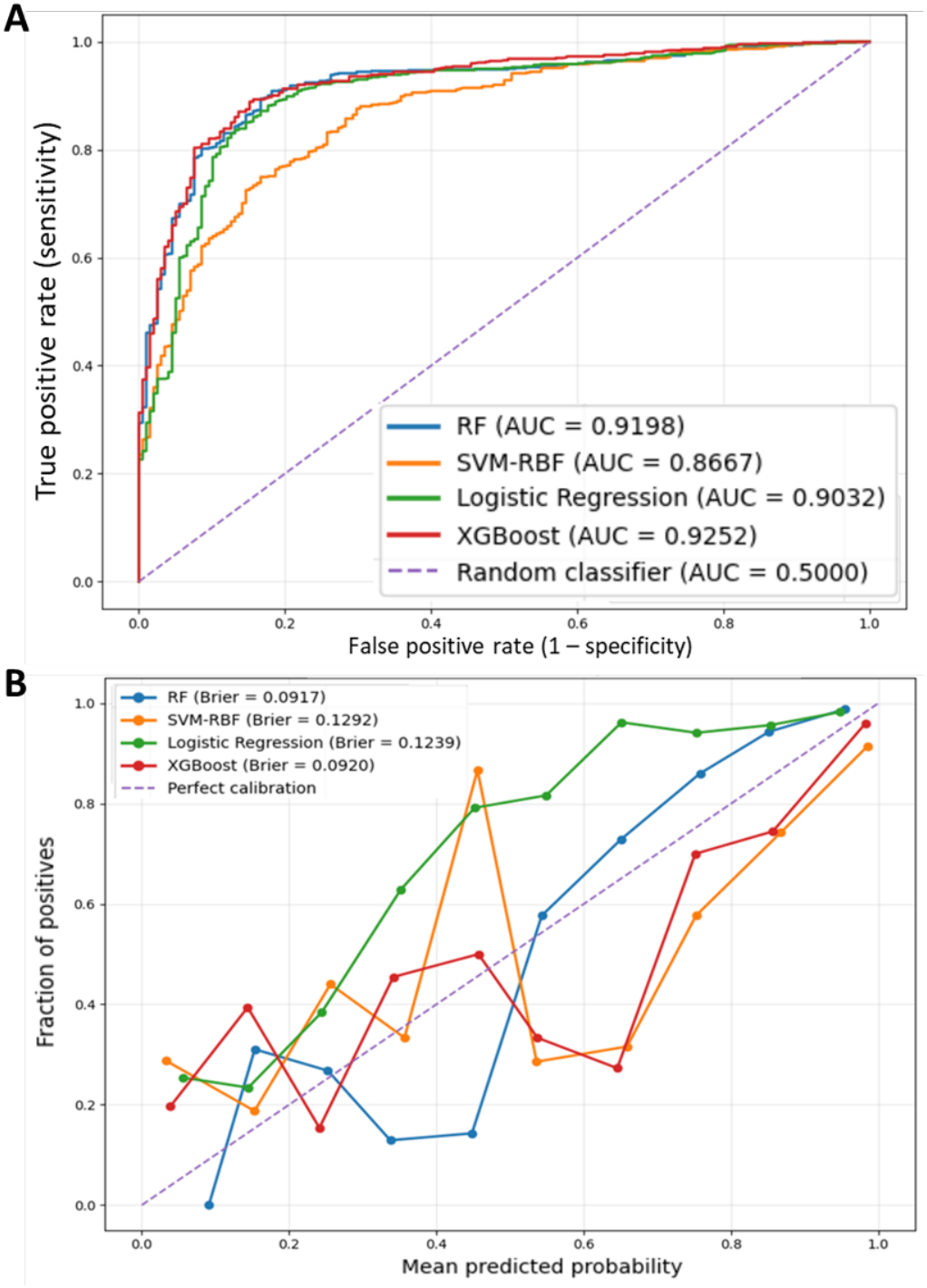


Figure 4. ROC and calibration curves of machine learning models. (A) ROC curves of the machine learning models on the independent test set. ROC curves comparing RF, SVM-RBF, LR, and XGBoost based on their ability to discriminate active and inactive S1PR1 compounds. (B) Calibration curves of the machine learning models on the independent test set. Comparison of predicted probabilities with observed outcome frequencies for RF, SVM-RBF, LR, and XGBoost. The diagonal line represents perfect calibration.

Abbreviations: LR: Logistic regression; RBF: Radial basis function; RF: Random forest; S1PR1: Sphingosine-1-phosphate receptor 1; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

subsequently used as a screening criterion to prioritize compounds from the external DrugBank dataset for downstream filtering and docking.

3.5. DrugBank virtual screening

The optimized RF model was subsequently used to score all 12,289 DrugBank compounds. The distribution of RF-derived activity scores showed a broad continuous distribution across the DrugBank library, with most compounds receiving scores below the selected screening threshold. Using an operational RF score threshold of ≥ 0.70, 277 of the 12,289 DrugBank compounds were retained as predicted active candidates for downstream analysis (Figure 5).

3.6. Lipinski-based compound prioritization

The 277 RF-prioritized DrugBank compounds with predicted activity probabilities ≥ 0.7 were subjected to an initial drug-likeness assessment using physicochemical properties obtained from the pkCSM web server. A total of 56 compounds passed all four Lipinski criteria and were retained for downstream ligand preparation for molecular docking.

3.7. Redocking validation and molecular docking

Redocking of the crystallographic JEU ligand was used to assess whether the docking protocol could reproduce the experimentally observed ligand pose in the S1PR1 binding site. At exhaustiveness values of 8, 16, and 32, the best predicted docking scores were −9.143, −9.542, and −9.542 kcal/mol, respectively. The corresponding RMSD values of the top-ranked poses were 1.957, 1.049, and 1.049 Å, respectively (Table 4. The top-ranked pose obtained at exhaustiveness 16 reproduced the crystallographic JEU conformation with an RMSD of 1.049 Å, while increasing exhaustiveness from 16 to 32 did not further change either the best docking score or the RMSD. Therefore, an exhaustiveness value of 16 was selected for subsequent compound docking.

3.8. Docking of prioritized compounds

A total of 56 compounds that had passed the preceding RF-based VS and Lipinski filtering steps were subjected to molecular docking against the prepared S1PR1 structure. The same binding-site coordinates and docking parameters established during JEU redocking were retained to ensure consistency throughout the docking workflow.

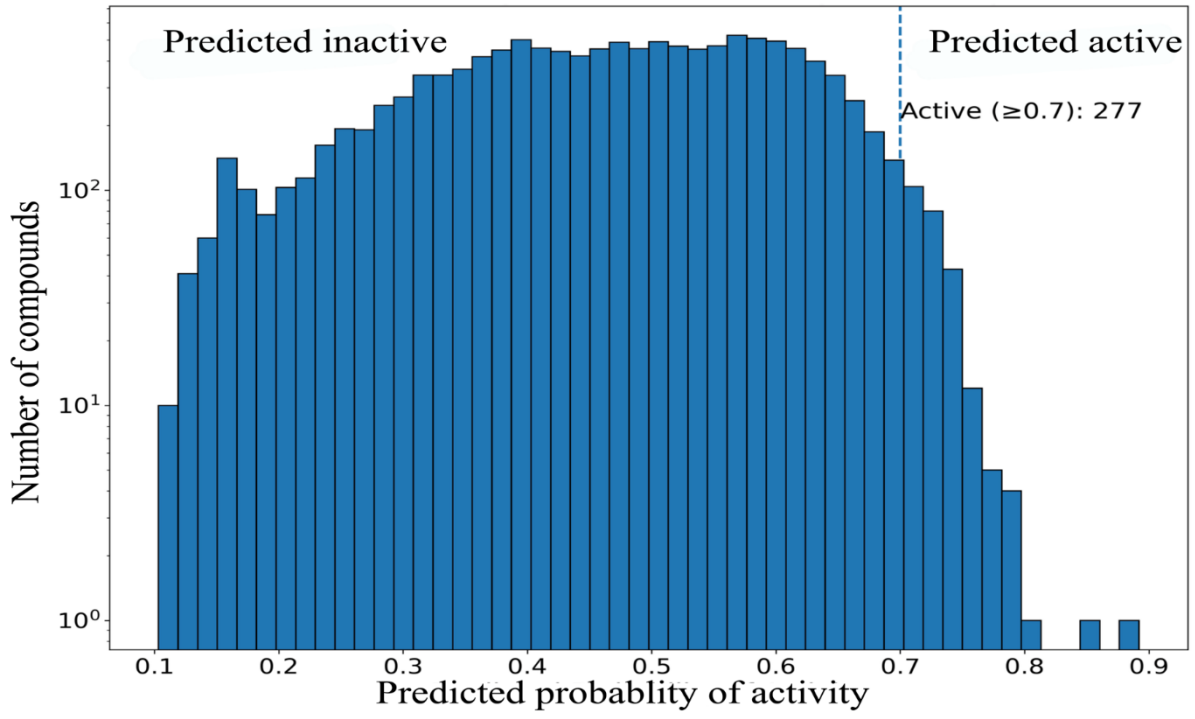


Figure 5. Distribution of random forest prediction probabilities for the DrugBank screening library. Distribution of predicted probabilities of activity for the 12,289 DrugBank compounds screened using the optimized random forest model. The dashed vertical line indicates the probability threshold of 0.70, above which 277 compounds were prioritized for subsequent virtual screening.

Table 4. Ozanimod redocking validation across docking exhaustiveness levels

Exhaustiveness	Binding affinity (kcal/mol)	RMSD (Å)
8	−9.143	1.957
16	−9.542	1.049
32	−9.542	1.049

Abbreviation: RMSD: Root-mean-square deviation.

Out of 56 compounds, 54 yielded successful docking results, whereas two compounds returned AutoDock Vina errors and were excluded from subsequent analysis. The successfully docked compounds showed predicted Vina scores ranging from −10.943 to −5.207 kcal/mol. A docking-score threshold of ≤ −7.5 kcal/mol was initially applied for computational prioritization, yielding 31 compounds.

Because detailed receptor–ligand interaction and ADMET characterization of all 31 compounds would provide a broad and less focused assessment, a further prioritization step was applied. The compounds were ranked by considering both the RF-predicted probability of activity and the predicted docking score, rather than relying on docking score alone. Based on these combined computational criteria, the top five compounds were selected for detailed structural interaction analysis and subsequent ADMET evaluation.

The docking scores were interpreted as computational scoring outputs for relative prioritization and not as direct measurements of experimental binding affinity. The subsequent interaction and ADMET analyses were therefore focused on the five prioritized compounds to identify candidates with favorable predicted S1PR1 binding characteristics and pharmacokinetic/toxicity profiles.

3.9. Protein–ligand interaction analysis

The crystallographic ozanimod binding environment was first examined to define the experimentally observed ligand-binding region. The ozanimod-associated binding pocket was operationally defined as the set of Chain D residues located within 4 Å of the crystallographic ozanimod ligand. The resulting pocket comprised residues including LYS34, ASN101, GLU121, MET124, PHE125, LEU128, SER129, VAL132, VAL194, LEU195, CYS206, THR207, VAL209, PHE210, LEU213, TRP269, LEU272, LEU276, and LEU297 (Figure 6). This experimentally defined pocket was subsequently used as a structural reference for evaluating the docking poses of the prioritized compounds.

The five prioritized compounds—udifitimod, AZD8529, tedizolid phosphate, mocravimod, and BMS-986104—were examined for their interactions within the ozanimod-associated S1PR1 binding pocket. The crystallographic ozanimod structure established a predominantly hydrophobic pocket involving residues such as Phe125, Leu128, Leu195, Cys206, Val209, Phe210, Leu213, Leu272, Leu297, together with polar residues including Ser129, Thr207, and Glu121.

The selected compounds occupied the same general binding cavity and established multiple contacts with residues forming the known ozanimod-associated pocket, although their individual interaction patterns differed (Figure 7).

Overall, the interaction analysis showed that the five prioritized compounds were positioned within the ozanimod-associated S1PR1 pocket and established multiple contacts with pocket-forming residues. The results provide structural support for their computational prioritization; however, these docking-derived interactions

Table 5. Physicochemical properties and drug-likeness of computationally prioritized S1PR1 candidates

PubChem CID	118877516	25125217	11476460	11155874	77050638	52938427
Compound	Udifitimod	AZD8529	Tedizolid phosphate	Mocravimod	BMS-986104	Ozanimod (JEU)
ML probability	0.892	0.784	0.787	0.751	0.848	Reference
Docking affinity (kcal/mol)	−10.943	−10.677	−10.125	−9.189	−8.772	−9.542
MW (Da)	379.544	487.482	450.323	443.996	329.528	404.47
LogP	4.39	3.505	1.513	4.685	4.719	3.632
Rotatable bonds	6	6	6	10	7	7
HBA	3	7	9	5	2	7
HBD	2	1	2	3	2	2
Surface area (Å^2)	168.469	199.21	174.02	186.204	147.393	174.859

Abbreviations: HBA: Hydrogen-bond acceptors; HBD: Hydrogen-bond donors; LogP: Lipophilicity; S1PR1: Sphingosine-1-phosphate receptor 1.

Table 6. Predicted ADMET properties of computationally prioritized S1PR1 candidates and ozanimod

PubChem CID	118877516	25125217	11476460	11155874	77050638	52938427
Compound	Udifitimod	AZD8529	Tedizolid phosphate	Mocravimod	BMS-986104	Ozanimod (JEU)
Water solubility (log mol/L)	−4.928	−5.567	−5.035	−5.558	−4.760	−3.154
Caco-2 permeability (log Papp)	1.267	1.065	0.015	0.946	1.135	0.637
Intestinal absorption (%)	95.089	91.717	75.259	91.773	92.434	91.053
P-gp substrate	Yes	Yes	Yes	Yes	Yes	Yes
VDss (log L/kg)	0.839	0.21	−0.469	0.631	0.851	0.628
Fraction unbound	0.068	0.301	0.386	0.304	0.326	0.16
BBB permeability (log BB)	0.081	−1.056	−1.662	−0.491	0.143	−0.606
CNS permeability (log PS)	−1.882	−2.566	−3.144	−2.529	−2.735	−2.562
CYP2D6 substrate	No	No	No	No	No	No
CYP3A4 substrate	Yes	Yes	Yes	Yes	Yes	Yes
CYP2D6 inhibitor	No	No	No	No	No	No
CYP3A4 inhibitor	Yes	Yes	No	Yes	No	Yes
Total clearance (log mL/min/kg)	1.316	0.449	0.023	0.985	1.397	0.69
OCT2 substrate	No	No	No	No	No	Yes
AMES	No	No	Yes	No	No	No
hERG I	No	No	No	No	No	No
hERG II	Yes	Yes	Yes	Yes	Yes	Yes
Hepatotoxicity	Yes	Yes	Yes	No	No	Yes

Abbreviations: ADMET: Absorption, distribution, metabolism, excretion, and toxicity; BB: Brain-to-blood ratio; BBB: Blood–brain barrier; CNS: Central nervous system; CYP: Cytochrome P450; hERG: Human ether-à-go-go-related gene; OCT2: Organic cation transporter 2; P-gp: P-glycoprotein; PS: Permeability–surface area product; S1PR1: Sphingosine-1-phosphate receptor 1; VDss: Volume of distribution.

are interpreted as predicted binding poses and not as experimental evidence of S1PR1 binding or modulation.

3.10. Lipinski drug-likeness and physicochemical properties

The five prioritized compounds showed physicochemical characteristics compatible with the principal Lipinski criteria. MWs ranged from 329.53 to 487.48 Da, while logP values ranged from 1.51 to 4.72. All candidates had ≤ 5 HBD and ≤ 10 HBA. Thus, none of the five candidates exceeded the four principal Lipinski thresholds. The reference ligand ozanimod (JEU) also remained within these criteria (Table 5).

3.11. ADMET prediction

The pkCSM analysis showed that all five candidates had predicted human intestinal absorption between 75.26% and 95.09%. Udifitimod showed the highest predicted intestinal absorption (95.09%), followed by BMS-986104 (92.43%), mocravimod (91.77%), and AZD8529 (91.72%). Tedizolid phosphate showed the lowest predicted absorption among the selected candidates (75.26%). Predicted Caco-2 permeability varied considerably, with udifitimod showing the highest value (1.267) and tedizolid phosphate the lowest (0.015) (Table 6).

All five candidates were predicted to be P-glycoprotein substrates, indicating that transporter-mediated disposition may contribute to their pharmacokinetic behavior. The predicted distribution parameters also varied among the compounds, including differences in volume of distribution, fraction unbound, and blood–brain barrier/CNS permeability.

With respect to metabolism, all five compounds were predicted to be CYP3A4 substrates, whereas none were predicted to be CYP2D6 substrates. CYP3A4 inhibition was predicted for udifitimod, AZD8529, and mocravimod, while CYP2D6 inhibition was not predicted for any candidate.

For toxicity-related endpoints, udifitimod, AZD8529, and tedizolid phosphate were predicted to show hepatotoxicity, whereas mocravimod and BMS-986104 were predicted as non-hepatotoxic. Tedizolid phosphate was also predicted to be AMES-positive, whereas the other

four candidates were AMES-negative. All five candidates were predicted as human ether-à-go-go-related gene II inhibitors, representing a potential cardiotoxicity-related liability that requires experimental investigation.

Importantly, the ADMET analysis was used as a complementary prioritization step rather than as evidence of experimental safety or efficacy. Together with ML prediction probability, docking affinity, Lipinski properties, and protein–ligand interaction analysis, the ADMET profiles supported the selection of udifitimod, AZD8529, tedizolid phosphate, mocravimod, and BMS-986104 as computationally prioritized S1PR1 candidates for further experimental investigation.

3.12. Chemical-space similarity and applicability-domain assessment

The chemical-space analysis revealed a substantial difference between the external DrugBank screening library and the independent ChEMBL test set with respect to their structural similarity to the ChEMBL training set.

For the 12,289 DrugBank compounds, the maximum Tanimoto similarity to the ChEMBL training set had a mean of 0.2480, a median of 0.2472, and a standard deviation of 0.0687. The observed values ranged from 0.0000 to 0.9825, with the 25th and 75th percentiles at 0.2105 and 0.2838, respectively. Only 2,179 compounds (17.73%) showed a maximum Tanimoto similarity ≥ 0.3, while 241 (1.96%) reached ≥ 0.4. Only 55 compounds (0.45%) had similarity ≥ 0.5, 17 (0.14%) had similarity ≥ 0.6, and 5 (0.04%) had similarity ≥ 0.7 (Table 7).

In comparison, the 901-compound ChEMBL test set showed substantially greater structural similarity to the training set. The mean and median maximum Tanimoto similarities were 0.6865 and 0.6957, respectively, with a standard deviation of 0.1367 and a range of 0.1910–1.0000. The 25th and 75th percentiles were 0.6164 and 0.7733, respectively. A total of 879 compounds (97.56%) had a maximum similarity ≥ 0.3, 871 (96.67%) ≥ 0.4, 828 (91.90%) ≥ 0.5, 712 (79.02%) ≥ 0.6, and 443 (49.17%) ≥ 0.7 (Table 7).

Overall, the DrugBank screening library exhibited markedly lower structural similarity to the ChEMBL training set than the independent ChEMBL test set, indicating that a large proportion of the external screening library occupied chemical space relatively distant from the model-development data. Thus, predictions for these compounds represent a greater degree of chemical-space extrapolation and were interpreted as computational prioritization rather than definitive experimental activity predictions.

4. Future perspectives

Future studies should experimentally validate the prioritized compounds using S1PR1 binding and functional assays to determine their actual receptor affinity, selectivity, and modulatory activity. Molecular dynamics simulations could subsequently be performed for the prioritized protein–ligand complexes to investigate the temporal stability of the predicted binding modes, receptor–ligand interactions, and conformational behavior of S1PR1 under dynamic conditions. Binding free-energy calculations, such as molecular mechanics Poisson–Boltzmann surface area or molecular mechanics

Table 7. Chemical-space similarity of the ChEMBL test set and DrugBank screening library to the ChEMBL training set

Similarity measure	ChEMBL test set (*n* = 901)	DrugBank screening library (*n* = 12,289)
Mean maximum Tanimoto	0.6865	0.2480
Median maximum Tanimoto	0.6957	0.2472
Standard deviation	0.1367	0.0687
Minimum	0.1910	0.0000
25th percentile	0.6164	0.2105
75th percentile	0.7733	0.2838
Maximum	1.0000	0.9825
≥0.3	879 (97.56%)	2,179 (17.73%)
≥0.4	871 (96.67%)	241 (1.96%)
≥0.5	828 (91.90%)	55 (0.45%)
≥0.6	712 (79.02%)	17 (0.14%)
≥0.7	443 (49.17%)	5 (0.04%)

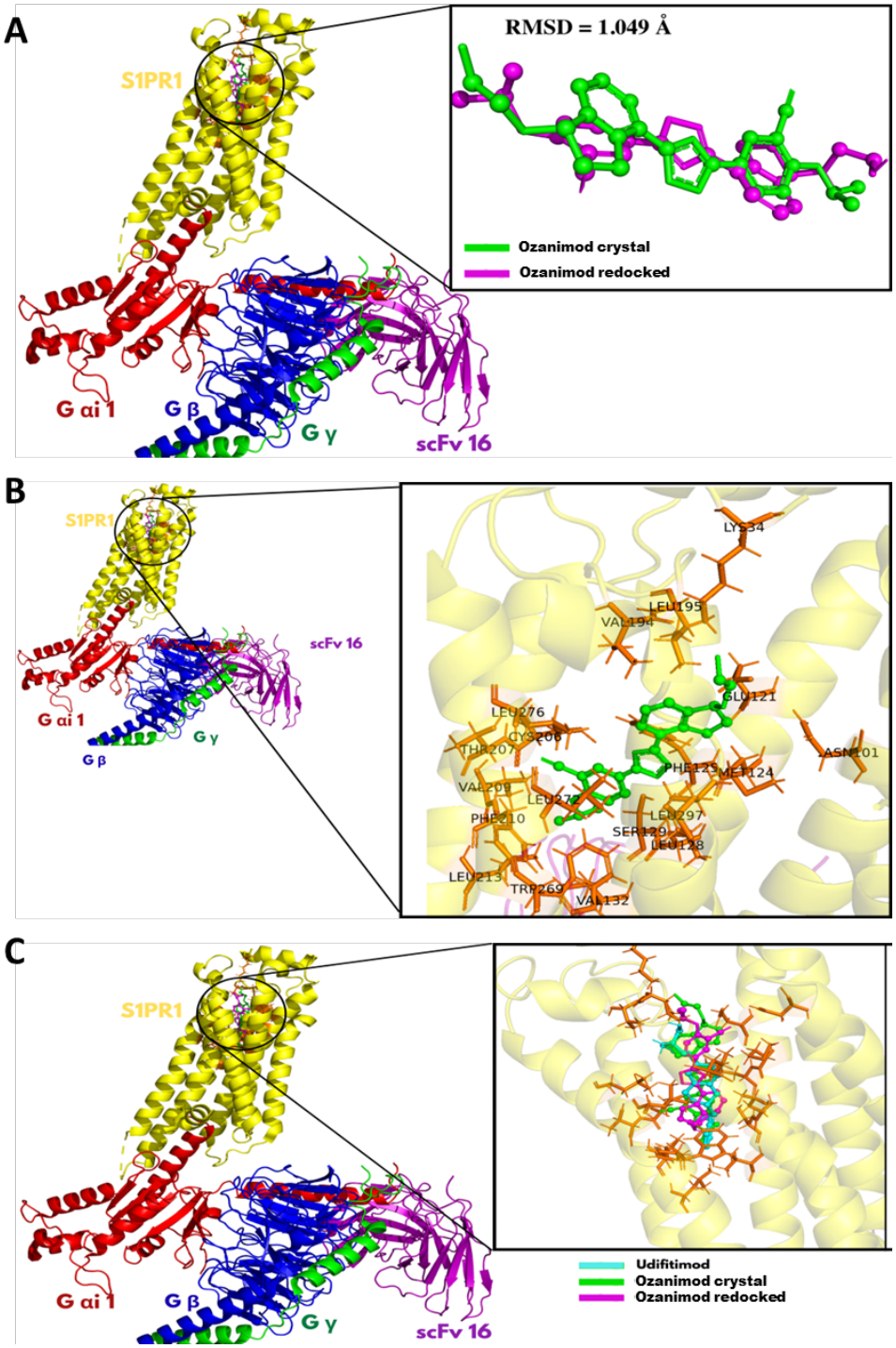


Figure 6. Structural validation of the S1PR1 docking protocol and predicted binding interactions of S1PR1 ligands. (A) Structural validation of the S1PR1 docking protocol by redocking the co-crystallized ligand ozanimod. Superposition of the crystallographic and redocked ozanimod poses within the S1PR1 binding site, showing an RMSD of 1.049 Å. (B) Binding-site interactions of ozanimod within the S1PR1 binding pocket. Three-dimensional representation of the S1PR1–ozanimod complex showing the ligand orientation and surrounding binding-site residues. (C) Predicted binding mode of udifitimod within the S1PR1 binding pocket. Three-dimensional representation of udifitimod positioned within the S1PR1 ligand-binding site, highlighting its predicted binding orientation and interactions with surrounding residues.
Abbreviations: RMSD: Root-mean-square deviation; S1PR1: Sphingosine-1-phosphate receptor 1.

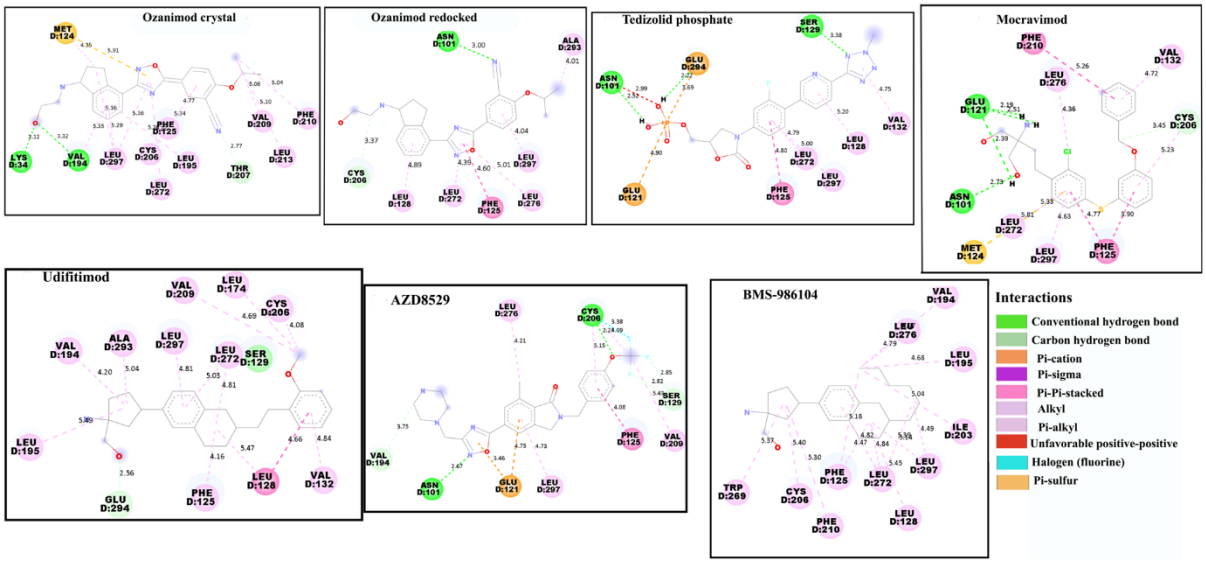


Figure 7. Two-dimensional interaction profiles of the reference ligand and prioritized S1PR1 candidates. Predicted binding interactions of crystallographic ozanimod, redocked ozanimod, and the five computationally prioritized compounds—udifitimod, AZD8529, tedizolid phosphate, mocravimod, and BMS-986104—within the S1PR1 binding pocket. The diagrams illustrate hydrogen-bond, hydrophobic, π-related, halogen, and other ligand–residue interactions identified from the docking poses.
Abbreviation: S1PR1: Sphingosine-1-phosphate receptor 1.

generalized Born surface area, could further be explored as complementary computational approaches for comparing the relative stability of the prioritized complexes. Additional studies could evaluate receptor subtype selectivity across S1PR1–S1PR5, followed by cellular and pharmacological assays to determine functional consequences of receptor engagement. Experimental ADMET and pharmacokinetic studies would also be required to verify the computational predictions. Finally, prospective screening of chemically diverse libraries followed by experimental testing could determine whether the proposed workflow can identify additional S1PR1-binding compounds beyond the currently prioritized candidates.

5. Conclusion

This study established an integrated machine-learning and structure-based virtual-screening workflow for the identification and computational prioritization of compounds with putative activity toward S1PR1. The machine-learning models demonstrated good discriminatory performance on an independent scaffold-separated test set, with RF and XGBoost showing the strongest overall predictive performance. Application of the optimized RF model to the external DrugBank library, followed by Lipinski filtering, molecular docking, protein–ligand interaction analysis, and ADMET assessment, resulted in the prioritization of five compounds: udifitimod, AZD8529, tedizolid phosphate, mocravimod, and BMS-986104.

Importantly, the identification of compounds with previously reported S1PR1-related pharmacological activity[55,56], including mocravimod and BMS-986104, provides supportive evidence that the computational workflow can prioritize chemically relevant S1PR1-targeting compounds. However, these observations are considered retrospective supportive evidence and not experimental validation of the present predictions. The docking results represent predicted binding poses and scoring outputs, while the ADMET results are computational predictions and do not establish experimental pharmacological activity, efficacy, or safety. Furthermore, chemical-space analysis showed that a substantial proportion of the DrugBank library was structurally distant from the ChEMBL training set, highlighting the extrapolative nature of many external-library predictions. Overall, the workflow provides a systematic computational strategy for prioritizing S1PR1-binding candidates and generating testable hypotheses for subsequent experimental investigation.

Acknowledgments

The authors gratefully acknowledge BioTecNika Info Labs Pvt. Ltd. for providing project-related assistance.

Funding

None.

Conflict of interest

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

Author contributions

Conceptualization: Akshata P. Pauskar
Data curation: Akshata P. Pauskar
Formal analysis: Akshata P. Pauskar
Investigation: Akshata P. Pauskar
Methodology: Akshata P. Pauskar
Project administration: Elamathi Natarajan
Supervision: Elamathi Natarajan
Writing – original draft: Akshata P. Pauskar
Writing – review & editing: Akshata P. Pauskar

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The datasets, computational workflows, molecular docking files, ADMET screening results, processed data, and other supplementary materials generated and analyzed in this study are publicly available through the Figshare repository at "doi:10.6084/m9.figshare.34103454". The repository includes the "Supplementary Files S1PR1.zip" archive containing the supplementary computational files associated with this study.

AI tools statement

During the preparation of this manuscript, ChatGPT (OpenAI, USA) was used solely to assist with improving the language, readability, and clarity of the text. All content generated or modified with the assistance of this tool was subsequently reviewed, edited, and verified by the authors. The authors take full responsibility for the accuracy, integrity, and final content of the published article.

References

1. Bravo GÁ, Cedeño RR, Casadevall MP, Ramió-Torrentà L. Sphingosine-1-Phosphate (S1P) and S1P Signaling Pathway Modulators, from Current Insights to Future Perspectives. *Cells.* 2022;11(13):2058.

   doi: 10.3390/cells11132058

2. Chaudhry BZ, Cohen JA, Conway DS. Sphingosine 1-Phosphate Receptor Modulators for the Treatment of Multiple Sclerosis. *Neurotherapeutics.* 2017;14(4):859–873.

   doi: 10.1007/s13311-017-0565-4

3. Chatzikonstantinou S, Poulidou V, Arnaoutoglou M, et al. Signaling through the S1P–S1PR axis in the gut, the immune and the central nervous system in multiple sclerosis: Implication for pathogenesis and treatment. *Cells.* 2021;10(11):3217.

   doi: 10.3390/cells10113217

4. Brunkhorst R, Vutukuri R, Pfeilschifter W. Fingolimod for the treatment of neurological diseases-state of play and future perspectives. *Front Cell Neurosci.* 2014;8:283.

   doi: 10.3389/fncel.2014.00283

5. Masuda-Kuroki K, Di Nardo A. Sphingosine 1-Phosphate Signaling at the Skin Barrier Interface. *Biology.* 2022;11(6):809.

   doi: 10.3390/biology11060809

6. Hisano Y, Kobayashi N, Yamaguchi A, Nishi T. Mouse SPNS2 Functions as a Sphingosine-1-Phosphate Transporter in Vascular Endothelial Cells. *PLoS ONE.* 2012;7(6):e38941.

   doi: 10.1371/journal.pone.0038941

7. Li J, Huang Y, Zhang Y, et al. S1P/S1PR signaling pathway advancements in autoimmune diseases. *Biomol Biomed.* 2023;23(6):922–935.

   doi: 10.17305/bb.2023.9082

8. Mendelson K, Evans T, Hla T. Sphingosine 1-phosphate signaling. *Development.* 2014;141(1):5–9.

   doi: 10.1242/dev.094805

9. Sassoli C, Pierucci F, Zecchi-Orlandini S, Meacci E. Sphingosine 1-phosphate (S1P)/ S1P receptor signaling and mechanotransduction: Implications for intrinsic tissue repair/regeneration. *Int J Mol Sci.* 2019;20(22):5545.

   doi: 10.3390/ijms20225545

10. Gaidarov I, Komori HK, Stepniak DT, et al. Unique pharmacological properties of etrasimod among S1P receptor modulators. *FEBS Open Bio.* 2025;15(1):108–121.

    doi: 10.1002/2211-5463.13907

11. Chen H, Wang J, Zhang C, et al. Sphingosine 1-phosphate receptor, a new therapeutic direction in different diseases. *Biomed Pharmacother.* 2022;153:113341.

    doi: 10.1016/j.biopha.2022.113341

12. Chaudhry B, Cohen J, Conway D. Sphingosine 1-Phosphate Receptor Modulators for the Treatment of Multiple Sclerosis. *Neurotherapeutics.* 2017;14(4):859-873.

    doi: 10.1007/s13311-017-0565-4

13. Roy R, Alotaibi AA, Freedman MS. Sphingosine 1-Phosphate

Receptor Modulators for Multiple Sclerosis. *CNS Drugs.* 2021;35(4):385-402.

doi: 10.1007/s40263-021-00798-w

14. Pérez-Jeldres T, Alvarez-Lobos M, Rivera-Nieves J. Targeting Sphingosine-1-Phosphate Signaling in Immune-Mediated Diseases: Beyond Multiple Sclerosis. *Drugs.* 2021;81(9):985–1002.

doi: 10.1007/s40265-021-01528-8

15. Park SJ, Im DS. Sphingosine 1-phosphate receptor modulators and drug discovery. *Biomol Ther.* 2017;25(1):80–90.

doi: 10.4062/biomolther.2016.160

16. Choon XY, Yeo JH, White C, Sharma E, Samaan MA. The Current Sphingosine 1 Phosphate Receptor Modulators in the Management of Ulcerative Colitis. *J Clin Med.* 2025;14(10):3475..

doi: 10.3390/jcm14103475

17. Dong J, Wu C, Lu T, et al. Enhancing multifunctional drug screening via artificial intelligence. *Digit Discov.* 2025;4(8):2012–2024.

doi: 10.1039/d5dd00082c

18. Marinho EM, Batista de Andrade Neto J, Silva J, et al. Virtual screening based on molecular docking of possible inhibitors of Covid-19 main protease. *Microb Pathog.* 2020;148:104365.

doi: 10.1016/j.micpath.2020.104365

19. Carpenter KA, Huang X. Machine Learning-based Virtual Screening and Its Applications to Alzheimer's Drug Discovery: A Review. *Curr Pharm Des.* 2018;24(28):3347–3358.

doi: 10.2174/1381612824666180607124038

20. Carlsson J, Luttens A. Structure-based virtual screening of vast chemical space as a starting point for drug discovery. *Curr Opinion Struct Biol.* 2024;87:102829.

doi: 10.1016/j.sbi.2024.102829

21. Roy S, Nagaraj K, Mittal A, Shah FC, Raja K. Artificial Intelligence in Virtual Screening: Transforming Drug Research and Discovery—A Review. *J Bio-X Res.* 2025:8.

doi: 10.34133/jbioxresearch.0041

22. Verma S, Kumar M, Khare N, Kumar Jha A, Prakash Rai N. Molecular Docking Study on the Effect on Lamin-B1 through Compounds for the Treatment of Multiple Sclerosis. *Int J Res Appl Sci Engin Technol.* 2021;9(VI):5052–5057.

doi: 10.22214/ijraset.2021.36121

23. Agu PC, Afiukwa CA, Orji OU, et al. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in disease management. *Sci Rep.* 2023;13(1):13398.

doi: 10.1038/s41598-023-40160-2

24. Hosseini M, Chen W, Xiao D, Wang C. Computational molecular docking and virtual screening revealed promising SARS-CoV-2 drugs. *Precis Clin Med.* 2021;4(1):1–16.

doi: 10.1093/pcmedi/pbab001

25. Alizadeh AA, Jafari B, Dastmalchi S. Drug Repurposing for Identification of S1P1 Agonists with Potential Application in Multiple Sclerosis Using In Silico Drug Design Approaches. *Adv Pharm Bull.* 2022;13(1):113–122.

doi: 10.34172/apb.2023.012

26. Helmke PS, Ecker GF. Refining Drug-Induced Cholestasis Prediction: An Explainable Consensus Model Integrating Chemical and Biological Fingerprints. *J Chem Inf Model.* 2025;65(11):5301–5316.

doi: 10.1021/acs.jcim.4c02363

27. Savander S, Curabaz NN, Abbasi AM, Khan A, Saeed K, Tanoli Z. Data-driven strategies for drug repurposing: insights, recommendations, and case studies. *Brief Bioinform.* 2025;26(6):bbaf625.

doi: 10.1093/bib/bbaf625

28. Burggraaff L, Lenselink EB, Jespers W, et al. Successive statistical and structure-based modeling to identify chemically novel kinase inhibitors. *J Chem Inf Model.* 2020;60(9):4283–4295.

doi: 10.1021/acs.jcim.9b01204

29. Dey V, Machiraju R, Ning X. Improving Compound Activity Classification via Deep Transfer and Representation Learning. *ACS Omega.* 2022;7(11):9465–9483.

doi: 10.1021/acsomega.1c06805

30. Turon G, Hlozek J, Woodland JG, Kumar A, Chibale K, Duran-Frigola M. First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. *Nat Commun.* 2023;14(1):5736.

doi: 10.1038/s41467-023-41512-2

31. Moshawih S, Bu ZH, Goh HP, et al. Consensus holistic virtual screening for drug discovery: a novel machine learning model approach. *J Cheminform.* 2024;16(1):62.

doi: 10.1186/s13321-024-00855-8

32. Sarkar C, Das B, Rawat VS, et al. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. *Int J Mol Sci.* 2023;24(3):2026.

doi: 10.3390/ijms24032026

33. Bian Y, Kwon JJ, Liu C, Margiotta E, Shekhar M, Gould AE. Target-driven machine learning-enabled virtual screening (TAME-VS) platform for early-stage hit identification. *Front Mol Biosci.* 2023;10:1163536.

doi: 10.3389/fmolb.2023.1163536

34. Machado LA, Krempser E, Guimarães ACR. A machine learning-based virtual screening for natural compounds

capable of inhibiting the HIV-1 integrase. *Front Drug Discov.* 2022;2:954911.

doi: 10.3389/fddsv.2022.954911

35. Visan AI, Negut I. Integrating Artificial Intelligence for Drug Discovery in the Context of Revolutionizing Drug Delivery. *Life.* 2024;14(2):233.

doi: 10.3390/life14020233

36. Yin L, Du X, Ma C, Gu H. Virtual Screening of Drug Proteins Based on the Prediction Classification Model of Imbalanced Data Mining. *Processes.* 2022;10(7):1420.

doi: 10.3390/pr10071420

37. Reddi Durgasree M, Sharma H, Kishen Ajay Kumar V, Jyothi V, Vinay Kumar S. AI-Driven Virtual Screening: Machine Learning-Based Prediction of Molecular Activity and Binding Affinity for Drug Discovery. *Int J Hum Comput Intell.* 2025;4(6):598–609.

doi: 10.5281/zenodo.17082841

38. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol.* 2010;63(8):938-939.

doi: 10.1016/j.jclinepi.2009.11.009

39. Kant S, Deepika, Roy S. Artificial intelligence in drug discovery and development: transforming challenges into opportunities. *Discov Pharm Sci.* 2025;1(1):7.

doi: 10.1007/s44395-025-00007-3.

40. Shrestha RL, Tamang A, Chhetri SP, et al. AI-assisted discovery of potent FGFR1 inhibitors via virtual screening and in silico analysis. *PLOS ONE.* 2025;20(9):e0331837.

doi: 10.1371/journal.pone.0331837

41. Dulsat J, López-Nieto B, Estrada-Tejedor R, Borrell JI. Evaluation of Free Online ADMET Tools for Academic or Small Biotech Environments. *Molecules.* 2023;28(2):776.

doi: 10.3390/molecules28020776

42. Eswaramoorthy R, Hailekiros H, Kedir F, Endale M. In silico molecular docking, DFT analysis and ADMET studies of carbazole alkaloid and coumarins from roots of *Clausena anisata*: a potent inhibitor for quorum sensing. *Adv Appl Bioinform Chem.* 2021;14:13–24.

doi: 10.2147/AABC.S290912

43. Umesh HR, Ramesh KV, Devaraju KS. Molecular docking studies of phytochemicals against trehalose–6–phosphate phosphatases of pathogenic microbes. *Beni Suef Univ J Basic Appl Sci.* 2020;9(1):5.

doi: 10.1186/s43088-019-0028-6

44. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* 2021;61(8):3891–3898.

doi: 10.1021/acs.jcim.1c00203

45. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31(2):455–461.

doi: 10.1002/jcc.21334

46. Butt SS, Badshah Y, Shabbir M, Rafiq M. Molecular Docking Using Chimera and AutoDock Vina Software for Nonbioinformaticians. *JMIR Bioinform Biotech.* 2020;1(1):e14232.

doi: 10.2196/14232

47. Rauf MohdA, Zubair S, Azhar A. Ligand docking and binding site analysis with PyMOL and autodock/vina. *Int J Basic Appl Sci.* 2015;4(2):168–177.

doi: 10.14419/ijbas.v4i2.4123

48. Azmal M, Paul JK, Prima FS, Talukder OF, Ghosh A, Buday L. An in silico molecular docking and simulation study to identify potential anticancer phytochemicals targeting the RAS signaling pathway. *PLOS ONE.* 2024;19(9):e0310637.

doi: 10.1371/journal.pone.0310637

49. Agrani W, Perera U, Mudalige H, Perera O. Site-specific docking against HIV-RT protein using AutoDock 4.2.6. GARI Int J Multidiscip Res. 2022;8(3):91-108.

50. Sharma V, Sharma PC, Kumar V. In Silico Molecular Docking Analysis of Natural Pyridoacridines as Anticancer Agents. *Adv Chem.* 2016;2016:1–9.

doi: 10.1155/2016/5409387

51. Rudrapal M, Chetia D. Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme. *J Drug Deliv Ther.* 2020;10(4):225–233.

doi: 10.22270/jddt.v10i4.4218

52. Pattnaik M, Banerjee P. Virtual Screening of Phytochemicals against NS5 Virus Protein and STAT2 Host Receptor: An In Silico Therapeutic Approach. *Int J All Res Educ Sci Methods.* 2024;12(9):970-978.

53. Dunn TB, López-López E, Kim TD, Medina-Franco JL, Miranda-Quintana RA. Exploring activity landscapes with extended similarity: is Tanimoto enough? *Mol Inform.* 2023;42(7):2300056.

doi: 10.1002/minf.202300056.

54. Fechner N, Jahn A, Hinselmann G, Zell A. Estimation of the applicability domain of kernel-based machine learning models for virtual screening. *J Cheminform.* 2010;2(1):2.

doi: 10.1186/1758-2946-2-2

55. Dhar TGM, Xiao HY, Xie J, et al. Identification and Preclinical Pharmacology of BMS-986104: A Differentiated $S1P_1$ Receptor Modulator in Clinical Trials. *ACS Med Chem Lett.* 2016;7(3):283-288.

doi: 10.1021/acsmedchemlett.5b00448

56. Dertschnig S, Gergely P, Finke J, et al. Mocravimod, a Selective Sphingosine-1-Phosphate Receptor Modulator, in Allogeneic Hematopoietic Stem Cell Transplantation for Malignancy. *Transplant Cell Ther.* 2023;29(1):41.e1-41.e9.
doi: 10.1016/j.jtct.2022.10.029