



OPEN Machine learning-based virtual screening and density functional theory characterisation of natural inhibitors targeting mutant PBP2x in *Streptococcus pneumoniae*

Avani Panickar^{1,2}, Anand Manoharan³ & Sudha Ramaiah^{1,2}✉

Streptococcus pneumoniae (*S. pneumoniae*) has developed resistance to β -lactam antibiotics, largely due to mutations in penicillin-binding protein 2x (PBP2x), particularly within conserved motifs such as STMK and KSG. PBP2x mutations are frequently reported in multidrug-resistant pneumococcal strains associated with pneumonia, meningitis, and septicemia, especially in serotypes 19A, 19F, and 23F, showing reduced susceptibility to β -lactam antibiotics. These mutations in the PBP2x disrupt antibiotic binding and enzymatic functions, highlighting the need for alternative therapeutic strategies. This study focused on five clinically relevant PBP2x mutations (T338A/G/P and K547G/T) within its active site. A library of phytocompounds was screened using a machine learning model trained to identify antibacterial compounds. Top candidates were filtered based on ADMET properties, and their electronic characteristics were assessed using HOMO–LUMO analysis and electrostatic potential mapping, through density functional theory (DFT). Glucozaluzanin C, a phytochemical derived from *Elephantopus scaber*, emerged as a potential candidate. Molecular docking and dynamics simulations revealed strong binding affinity and structural integrity with all PBP2x mutants, over a 100-ns timescale. RMSD, RMSF, and hydrogen bonding analysis confirmed stable interactions, suggesting Glucozaluzanin C may effectively interact with PBP2x mutants. Overall, the study highlights an effective strategy for identifying plant-derived inhibitors against β -lactam-resistant *S. pneumoniae*.

Keywords Glucozaluzanin c, Penicillin binding protein 2x, Molecular dynamic simulation, Molecular docking, *Elephantopus scaber*, β -lactam-resistant

The prevalence of multidrug-resistant *Streptococcus pneumoniae* (*S. pneumoniae*) has been increasing since it was first reported in the 1960s¹. *Streptococcus pneumoniae*, a clinically important pathogen, causes a broad spectrum of invasive and non-invasive infections such as pneumonia, meningitis, septicemia, and otitis media. It remains a major contributor to morbidity and mortality, particularly among children under five years of age, adults over 60 years, and immunocompromised individuals. Although the widespread administration of β -lactam antibiotics has significantly contributed to the management of pneumococcal infections for decades, the emergence of resistance to these antibiotics has become a global concern. In some countries, *S. pneumoniae* strains now show reduced susceptibility to broad β -lactams such as amoxicillin^{2,3}. The mechanism of resistance in *S. pneumoniae* is largely recognised by the structural modification in penicillin-binding proteins (PBPs), which serve as the main targets of β -lactams. *S. pneumoniae* genome encodes six PBPs, and among these, PBP2x plays a major role in peptidoglycan cross-linking and cell wall biosynthesis and is one of the first PBPs to acquire mutations under antibiotic pressure⁴. Mutations in the PBP2x protein can result in low-affinity PBP2x variants, reducing the efficacy of the β -lactam antibiotics and contributing to antimicrobial resistance (AMR)⁵. Specific amino acid substitutions in the transpeptidase domain of PBP2x, such as T338A/G/P and K457G/T, have been identified for causing reduced β -lactam binding⁶. These mutations alter the geometric charge distribution of the active site, weakening the interaction between the antibiotic and the serine residue required for acylation and

¹Medical and Biological Computing Laboratory, School of Biosciences and Technology, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu 632014, India. ²Department of Bio-Sciences, School of Biosciences and Technology, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu 632014, India. ³Infectious Diseases Medical and Scientific Affairs, GlaxoSmithKline (GSK), Worli, Maharashtra 400030, India. ✉email: sudhaanand@vit.ac.in

inhibition^{7,8}. In clinical isolates, mutations are often associated with a high level of resistance, making traditional antibiotics ineffective. The T338 residue, part of the SXXX motif, plays a crucial role in forming hydrogen bond networks and stabilising substrate interactions. Alteration at this site leads to an “open” active site configuration, disrupting water-mediated interactions essential for drug binding⁹. Similarly, mutations at K547 near the catalytic groove affect the electrostatic interactions and substrate stabilisation, decreasing drug effectiveness. Understanding these structural changes highlights the importance of focusing on specific PBP2x mutations when designing more effective, targeted inhibitors¹⁰. Our previous work¹¹ identified high-frequency PBP2x mutations in clinical pneumococcal isolates recovered from patients with pneumonia, meningitis, and septicæmia in India and Vietnam. Similar patterns have been reported in China, South Korea, and Japan, particularly in multidrug-resistant serotypes like 19A, 19F, 6B, and 23F. These serotypes are commonly associated with both community-acquired pneumonia and invasive pneumococcal diseases and are frequently resistant to β -lactams. The regional spread and resistance profile underscore the importance of developing new antimicrobial inventions targeting conserved mutations in PBP2x¹².

As β -lactam antibiotics face increasing limitations against resistant PBP2x variants, the investigation for novel therapeutic strategies has been directed towards plant-driven natural compounds. Phytochemicals such as phenolics, flavonoids, and terpenoids exhibit diverse chemical scaffolds and have demonstrated intrinsic antibacterial properties. Some are known to bind and inhibit PBPs allosterically^{13,14}. Recent in silico studies have reported promising interactions of phytochemicals such as silicritin, galocatechin gallate, and curcumin analogues with PBP2x, suggesting their potential role in overcoming resistance mechanisms^{8,15}. Based on this evidence, Machine learning (ML) algorithms analyse large-scale molecular datasets to identify patterns in physicochemical properties and predict the possibility of strong binding to specific protein targets. By combining ML-driven prediction models with high-throughput virtual screening, it is possible to prioritise phytochemicals that exhibit favourable interaction profiles^{16,17}.

The present study builds on our previous work and aims to identify plant-derived inhibitors targeting mutant forms of PBP2x in *S. pneumoniae*, specifically those with T338A/G/P and K547G/T substitutions associated with β -lactam resistance. Accordingly, we utilised a screening approach that integrates ML-based compound screening, density functional theory (DFT) analysis for reactivity validation, and molecular docking simulations to evaluate binding interactions as represented in Fig. 1. By focusing on structurally characterised resistance mutations, this study aims to discover phytochemicals with high binding affinity and structural stability, contributing to the development of targeted therapeutic options against β -lactam-resistant *S. pneumoniae*.

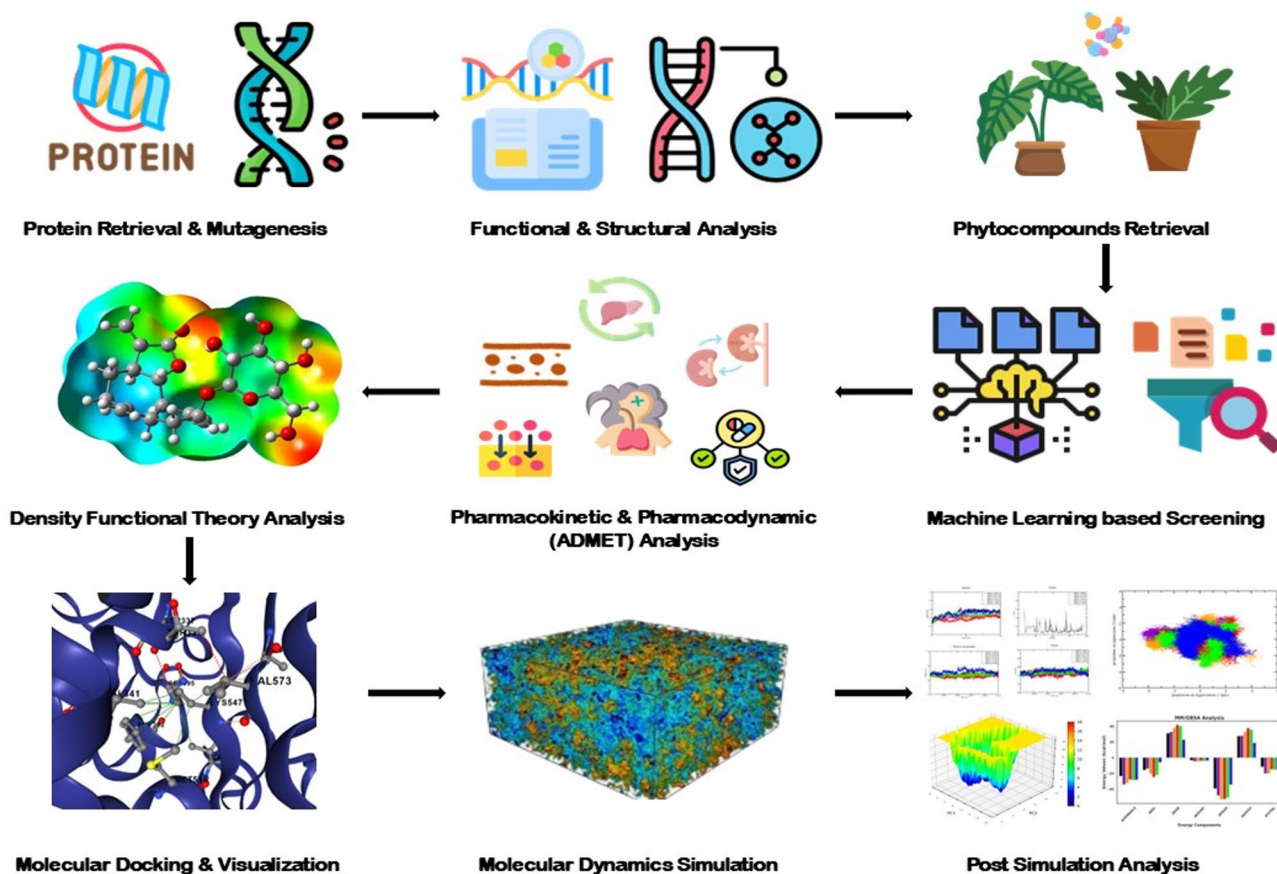


Fig. 1. Schematic workflow of machine learning-based virtual screening and DFT characterisation of natural inhibitors targeting mutant PBP2x in *S. pneumoniae*.

Methods

Protein retrieval and mutant generation

The reference sequence of the PBP2x protein (Accession ID: COS99621) from *S. pneumoniae* (GenBank Accession: LN831051) was retrieved from the NCBI database. A BLASTp search against the Protein Data Bank (PDB) was then conducted to identify a suitable 3D crystal structure. The structure with the highest sequence similarity (> 99%) and maximum alignment coverage was chosen, specifically the crystal structure of PBP2x (PDB ID: 1QME). The selected structure was visualised using PyMOL software, and all heteroatoms, including water molecules and ligands, were removed to prepare the protein for subsequent analysis. Following structural preparation, site-directed mutagenesis was carried out using the SPDB Viewer to introduce clinically relevant amino acid substitutions. The mutations T338A/G/P within the conserved SXXK (STMK) motif, and K547G/T adjacent to the KXG (KSG) motif were introduced. These mutations have been previously associated with reduced susceptibility to amoxicillin and are commonly observed in multidrug-resistant clinical isolates^{18,19}. The prepared wild-type and mutant protein models were then used for further analysis²⁰.

Evaluation of functional and structural impact of mutations

To assess the functional impact of amino acid substitutions in the PBP2x protein, we used PROVEAN and Meta-SNP. PROVEAN predicts whether an amino acid substitution or indel has a neutral or deleterious based on sequence homology and alignment-based scoring²¹. Meta-SNP, a consensus classifier integrating SIFT, PhD-SNP, SNAP, and PANTHER, was used for cross-validation to determine single-nucleotide polymorphism SNP pathogenicity²². We also applied ConSurf for identifying the evolutionarily conserved region within the PBP2x²³. Additionally, we used DynaMut2 to evaluate the impact of mutations on protein stability. It combines normal mode analysis and graph-based features to estimate $\Delta\Delta G$ values, indicating stabilizing and destabilizing effects²⁴.

Phytocompound curation and model prediction

Data preparation

Bioassay data for the compounds active against *S. pneumoniae* were retrieved from the PubChem Bioassay database (AID 438298), comprising 36 active and 5 inactive compounds. To improve classification performance and address class imbalance, additional inactive compounds were included. These curated compounds formed the core dataset for model development. For external testing, structurally complete phytocompounds with traditional use against pneumonia were selected from the IMPPAT (Indian Medicinal Plants, Phytochemistry and Therapeutics) database²⁵.

Descriptor calculations

All compounds were converted to 3D-Standard Format (3D-SDF), and molecular descriptors were generated using PaDEL-Descriptor^{26,27}. This included 1D, 2D, and 3D features along with chemical fingerprints such as MACCS keys and PubChem substructures. Feature selection was carried out using WEKA (Waikato Environment for Knowledge Analysis) built-in evaluation methods to retain only the most informative descriptors for model training^{28,29}.

Machine-learning model prediction

Supervised learning was performed using tenfold cross-validation³⁰. Classifiers including Random Forest, J48, PART, Reduced Pruning Tree (RepTree) were tested within WEKA, chosen for their proven reliability in virtual screening applications^{31,32}. Model performance was evaluated using standard matrices such as accuracy, precision, sensitivity, specificity, F1-score, and area under the ROC curve (AUC)³³. Confusion matrices were generated to assess classification outcomes and ensure robustness across models.

ADMET screening evaluation

The ADMET (absorption, distribution, metabolism, excretion, and Toxicity) profiling was carried out to evaluate the pharmacokinetic properties and safety of the selected phytocompounds, which are essential for assessing drug-likeness³⁴. The analysis used ADMETlab 3.0, a web-based tool that predicts multiple ADMET parameters that is widely used in virtual screening workflows for both natural and synthetic compounds^{35,36}. Toxicity was assessed using ProTox 3.0, which predicts endpoints such as LD50, hepatotoxicity, and carcinogenicity based on structure-activity relationships and machine learning models^{37,38}. The combined use of ADMETlab and ProTox enables efficient screening of the candidate phytocompounds and eliminates those with poor pharmacokinetic and safety profiles. Compounds showing favourable ADME properties and minimal toxicity were selected and used for further analysis.

Density functionality theory (DFT) analysis

The stability and reactivity of selected phytocompounds were evaluated using DFT, a quantum chemistry method widely employed to examine molecular structures, electronic properties, and reactivity³⁹. Geometry optimisations were performed using the B3LP functional with the 6-311g++ (d, p) basis set in Gaussian09W⁴⁰. structural visualisation was carried out in GaussView 6.0. To assess molecular reactivity, frontier molecular orbitals (FMOs) [Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbitals (LUMO)], were analysed. Key reactivity descriptors such as ionisation potential, electron affinity, chemical hardness, and electrophilicity were calculated based on Koopmans theorem. Electrostatic potential (ESP) maps were generated to visualise electron-rich and electron-deficient regions, aiding the identification of likely nucleophilic or electrophilic attack sites⁴¹. Frequency calculations confirmed that the optimised geometries corresponded to true energy minima (no imaginary frequencies). Additionally, IR and Raman spectra were simulated to support structural characterisation⁴².

Protein-ligand interactions analysis

Molecular docking analyses were conducted to explore the interactions between the selected compound and the wildtype as well as mutant forms of PBP2x. Docking was performed using AutoDock v1.5.7, a widely used and validated platform for structure-based drug design^{43,44}. Both the wild type and mutant structures of PBP2x were prepared for docking by introducing polar hydrogen atoms, Kollam unit atom charges and merging non-polar hydrogen atoms. This preparation step stabilises the finishing point and ensures accurate representation of the protein's electrostatic surface⁴⁵. Ligands were prepared by fixing torsions and assigning Gasteiger charges, allowing for precise modelling of conformational flexibility and charge distribution⁴⁶. The active site of PBP2x, including mutation hotspots (T338A/G/P and K547G/T), was visualised using PyMOL 2.5. Coordinates for the docking grid were extracted based on the known binding pocket, centred at X = 110.427, Y = 62.793, and Z = 74.090, with grid dimensions of 60 × 60 Å³ and a spacing of 0.375 Å, ensuring complete coverage of the active site⁴⁷. Docking analyses were performed using the Lamarckian Genetic Algorithm, which incorporates evolutionary search with local optimisation for efficient conformation generation. The lowest binding energy (BE) docked structure was selected for each protein-ligand pair and further analysed⁴⁸. The best docked complexes were visualised and analysed using BIOVIA Discovery Studio Visualizer 2020⁴⁹. The best complex was further used for the Molecular Dynamics (MD) simulation.

Stability assessments of Protein-Ligand complex

MD simulations were carried out using GROMACS v2024.2⁵⁰, to evaluate the stability and dynamics of the protein-ligand complexes. Docked complexes involving the PBP2x wild-type and mutant forms were selected, along with the apo form of the protein for comparative analysis⁵¹. System topologies were generated using the CHARMM general field of force (CGenFF)⁵², which provides parameters suitable for drug-like molecules. Each system was placed in a dodecahedral simulation box with a minimum 1.0 nm distance from the solute and solvated using TIP3P water molecules⁵³. Counter-ions were added to neutralise the system and replicate physiological ionic strength. Energy minimisation was performed using the steepest descent algorithm for 50,000 steps, with a force convergence criterion of 10 kJ/mol. The Verlet cutoff scheme was used to handle non-bonded interactions⁵⁴. Equilibration was carried out in two phases: first under the NVT ensemble (100 ps) to stabilise temperature, followed by NPT equilibration (100 ps) to stabilise pressure, both at 310 K and 1 bar. The velocity-rescale thermostat was used for temperature coupling, and the Parrinello-Rahman barostat was applied for pressure coupling^{55,56}. A 100 ns production MD run was conducted with periodic boundary conditions and a 2 fs integration time step. Electrostatic interactions were treated using the Particle Mesh Ewald (PME) method, and all bond lengths involving hydrogen atoms were constrained using the LINCS algorithm.

Trajectory analyses were performed to evaluate structural stability and motions. Principal Component Analysis (PCA) was used to identify major conformational changes based on the backbone atom fluctuations⁵⁷. The covariance matrix was calculated using *gmx covar*, and eigenvectors were extracted via *gmx anaeig*. The first two principal components (PC1 and PC2), representing dominant motions^{58,59}, were used to generate Free Energy Landscape (FEL) plots. FELs were constructed using a Python script that computed energy basins and conformer stability across the simulation landscape⁶⁰. Biding free energy calculations were performed using the Molecular Mechanics/ Generalised Born Surface Area (MM/GBSA) incorporated in *gmx_MMPBSA* program. A total of 10,000 frames were extracted from the 100ns trajectory⁶¹. The solvation model parameter (*igb*) was set to 5, with internal and external dielectric constants of 1.0 and 78.5, respectively, to simulate polarisation effects⁶². In addition, per-residue decomposition analysis was performed to quantify individual amino acid contributions to the overall binding energy, identifying key interacting residues across complexes^{63,64}.

Result

Mutant and structure generation

The mutations selected for this study were identified through a comprehensive literature review, which highlighted clinically relevant mutants in the PBP2x, a region critical for β-Lactam antibiotic binding. Specifically, we focused on two key active-site motifs, mainly SXXX and KSG. Mutations at positions T338A/G/P (within the STMK motif) and K547G/T (within the KSG motif) were targeted. These mutations were introduced into the wild-type PBP2x structure retrieved from the PDB. A total of five mutations (T338A, T338G, T338P, K547G, K547T) were induced into the structure for further analysis. Notably, both mutation sites are located within the same binding pocket.

Mutation impact analysis

The functional impact of amino acid substitutions in the PBP2x protein was assessed using SNP tools showed that all five selected mutations that are STMK (T338A, T338G, T338P) and KSG (K547G, K547T) were predicted to be deleterious (Supplementary Table S1). Conservation analysis indicated that these residues are highly conserved and are positioned in functionally important regions of the protein. Surface accessibility mapping further showed that these residues are surface-exposed. Protein stability changes upon mutation were evaluated, which calculates ΔΔG values based on normal model analysis. All five mutations showed negative ΔΔG values, indicating a destabilising effect on protein structure. The extent of destabilisation varied among the mutations, as summarised in Supplementary Table S2. The predicted structural change for the protein in the Supplementary Figs. S1 and S2.

ML-Based screening of phytocompounds

A total of 10,000 phytocompounds were initially collected from 111 different plant species known for their activity against bacterial pneumonia. After screening and removing all the duplicate entries, a final dataset of 2658 compounds was obtained. To create the best ML model for the virtual screening of these phytocompounds,

we constructed different classification models using an ML approach. Six different classification models were developed. These included Discussion Stump, Random Forest, Random Tree, Rep Tree, J48 Tree, and PART rules. The models trained and validated by means of tenfold cross-validation on the training dataset to ensure reliability and generalisation. The performance of each model was evaluated using standard statistical metrics, as shown in the corresponding Supplementary Table S3. Among the models tested, the REP tree model showed the highest performance. It achieved the highest Kappa statistic value of 0.859, indicating a high level of agreement between the predicted classifications and the actual classes. A value of 1 in Kappa statistics represents perfect agreement, and the results from the REP tree model were close to this ideal. Additionally, this model also showed the lowest root mean square error value of 0.25, indicating fewer deviations between predicted and actual outcomes. The REP tree model also showed the highest overall accuracy, attaining 92.95%, surpassing all the other models. Furthermore, the sensitivity and specificity of the REP Tree model were found to be 97% and 88%, respectively, which were the highest amongst all the classifiers. In contrast, the Randomized Tree classifier model showed the lowest sensitivity and specificity values at 66% and 68%, respectively.

Further, to access the model, the Receiver Operating Characteristics (ROC) curve used for the evaluation method helps to visualise the performance of binary classification models at different discrimination thresholds. The ROC curve for the REP Tree model showed the strongest relationship between the true positive rate (sensitivity) and the false positive rate, thus indicating its superior classification ability. It recorded the highest ROC value among all models, while the Random Tree model exhibited the lowest ROC performance. Based on the overall performance in all evaluated parameters, such as accuracy, sensitivity, specificity, RMSE, Kappa statistics and ROC curve, the REP Tree model was selected for virtual screening of the phytocompounds as illustrated in Supplementary Fig. S3. Using this model, it was observed that out of the 2658 screened natural compounds, 745 were classified as active, while 1913 were identified as inactive. All 745 active compounds were selected for subsequent screening and further analysis.

Pharmacokinetic and pharmacodynamic profiling

After initial virtual screening, a total of 745 compounds were screened for ADMET profiles to narrow them down based on the key pharmacokinetic and toxicity parameters. The first step we evaluated the Lipinski Rule of five, checking molecular weight, logP, number of hydrogen bond acceptor and donors. Compounds violating more than one of these criteria are predicted to have poor absorption and permeability. Based on this, 744 compounds satisfied the Lipinski rule and were retained for further screening. Next, we evaluated Human Intestinal Absorption (HIA). Compounds with an absorption rate below 30% are considered poorly absorbed. Using a threshold of 0.4, a total of 324 compounds were found to meet the criteria. Following, we checked the interaction with P-glycoprotein (P-gp), a key membrane protein involved in drug transport. Compounds with P-gp substrate scores within a range of 0.7 were considered likely to avoid efflux, and 309 compounds satisfied this condition. We then assessed Plasma Protein Binding (PPB), as compounds with high protein binding (above 90%) may have limited therapeutic activity due to reduced free drug availability. Based on this, compounds with PPB values less than 90% were selected, reducing the total to 132 compounds. To ensure that the compounds could effectively reach the central nervous system, we screened for blood-brain barrier (BBB) permeability using a threshold of 0.7, identifying 126 compounds with potential BBB penetration, suggesting their potential to exhibit good treatment efficacy against bacterial meningitis. Subsequently, we analysed plasma clearance (CL_{plasma}), which is important for determining drug dosage and exposure. Compounds with a clearance rate around 0.5 were considered optimal, and their step filtered the number of compounds down to 67. Finally, we performed toxicity prediction focusing on compounds showing predicted Toxicity classes higher than 4 and having LD50 values above 2000 mg/kg. Compounds showing predicted toxicity in hepatotoxicity, carcinogenicity, respiratory and mutagenicity were considered. This step reduced the number of compounds to five that met all ADMET and toxicity criteria and were selected for further analysis (Supplementary Table S4).

Density functional theory analysis

The electronic and structural properties of the selected compounds are closely related to their frontier molecular orbitals (FMOs), specifically the HOMO and the LUMO. These orbitals play a crucial role in defining the electron-donating and electron-accepting capacities of ligands. Understanding the energy difference between HOMO and LUMO levels is an important approach for assessing the chemical reactivity, kinetic stability and overall molecular stability of the compounds. Normally, a smaller HOMO-LUMO energy gap is associated with higher chemical reactivity and increased potential for biological activity, whereas a larger energy gap is associated with higher molecular stability and reduced reactivity. The five compounds that were screened from the ADMET were used for the prediction of their molecular stability and to understand their kinetic stability and chemical reactivity based on their FOMO structures. The compounds evaluated were 1063; 3999; 4703; 5991 and 6552. The corresponding HOMO-LUMO energy gaps for these compounds were found to be 0.68878 eV, 4.5843 eV, 4.1978 eV, 4.6024 eV, 4.0354 eV, respectively. A lower energy gap is associated with strong chemical reactivity and potential biological activity, whereas a larger gap suggests a more stable and less reactive molecule. Based on the HOMO and LUMO energies, additional chemical parameters including Ionization Potential (IP), Electron Affinity (EA), Electronegativity (χ), chemical hardness (η), Chemical Softness (S) were calculated and are summarised in the Table 1. The HOMO and LUMO molecular orbital distributions, along with the molecular electrostatic potential (MEP) maps of the compound, is depicted in the Fig. 2a and b. In the MEP maps, blue colour indicates regions of significant positive electrostatic potential (electron-deficient areas), whereas the red colour indicates electron-rich regions. Electronegative atoms, such as oxygen and carbon, were observed in the red regions. Nucleophilic species preferentially target regions with higher negative electrostatic potential (red zones), while electrophilic species are attracted to areas with lower positive potential (blue zones). It was observed that the compound 1063 showed higher electronegativity and electrophilicity values compared to

Parameters (eV)	IMPHY001063	IMPHY003999	IMPHY004703	IMPHY005991	IMPHY006552
HOMO	− 5.71087	− 6.2074	− 6.0967	− 5.7147	− 5.8339
LUMO	− 0.022092	− 1.6231	− 1.8989	− 1.1123	− 1.7985
Energy Gap (ΔE)	5.689	4.584	4.198	4.602	4.035
Ionization potential (IP)	5.711	6.207	6.097	5.715	5.834
Electron affinity (EA)	0.022	1.623	1.899	1.112	1.799
Chemical potential (μ)	− 2.86648	− 3.91525	− 3.9978	− 3.4135	− 3.8162
Global hardness (η)	2.844	2.292	2.099	2.301	2.018
Softness (S) (eV^{-1})	0.352	0.436	0.476	0.434	0.496
Electronegativity (χ)	2.866	3.915	3.998	3.414	3.816
Electrophilicity index (ω)	1.445	3.343	3.805	2.531	3.610

Table 1. DFT- calculated electronic parameters of phytocompounds for PBP2x screening.

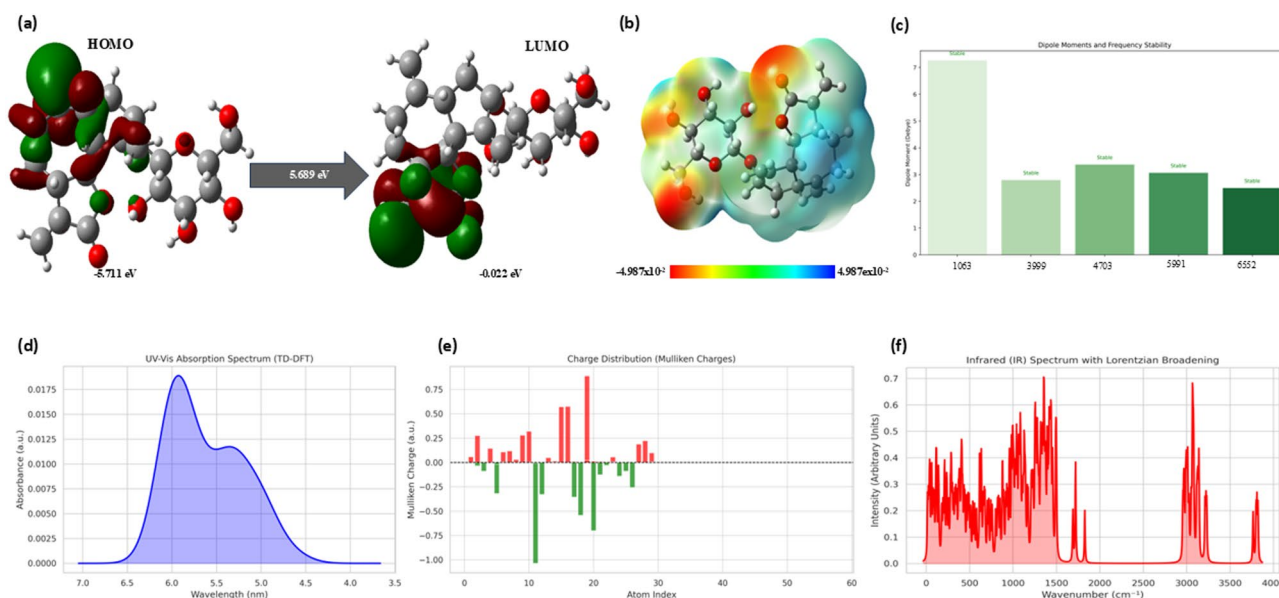


Fig. 2. Density functional theory (DFT) and time-dependent DFT (TD-DFT) analysis of glucosaluzanin C (a) Frontier molecular orbitals (FMOs) depicting the HOMO (− 5.71 eV) and LUMO (− 0.022 eV) distributions with an energy gap of 5.689 eV, indicating electronic excitation potential (b) Molecular Electrostatic Potential (MEP) surface showing charge distribution across the molecular surface; red and blue regions represent electron-rich and electron-deficient areas, respectively. (c) Dipole moment and vibrational frequency analysis indicating molecular stability across normal modes (values shown for selected modes) (d) UV-Vis absorption spectrum simulated using TD-DFT, showing maximum absorbance in the range of 6.0–6.5 nm (e) Mulliken charge distribution analysis across atomic indices, highlighting regions of electron donation (negative charge) and acceptance (positive charge). (f) Simulated Infrared (IR) spectrum with Lorentzian broadening, identifying key vibrational modes contributing to molecular characterization.

the other compounds, further supporting its potential for strong interactions and reactivity. The MEP analysis provided insights into the nucleophilic and electrophilic regions, with nucleophilic reactions associated with blue zones and electrophilic reactions linked to red zones. These observations were further supported by Time-Dependent Density Functional Theory (TD-DFT) based Ultraviolet-Visible (UV-Vis) absorption analysis, which indicated significant electronic transitions, Mulliken charge distribution highlighting polarised reactive centres, Infrared (IR) vibrational spectra confirming no imaginary frequencies and vibrational stability and the high dipole moment suggesting strong polarity and structural stability (Fig. 2c and f, Supplementary Figs. S4–S7).

Molecular docking and intermolecular interactions analysis

Molecular docking analysis was conducted to evaluate the binding potential of selected phytocompounds with the active site residues of the PBP2x protein. The active binding sites of PBP2x included two conserved motifs: STMK and KSG. The STMK site comprised the residues SER337, THR338, MET339, LYS340, whereas the KSG site included LYS547, SER548, and GLY549. These residues were selected based on their structural proximity to the ligand-binding pocket and their functional relevance to the catalytic activity of PBP2x. To assess the effect of mutations at these active sites, five mutants of PBP2x were used in docking analysis along with five

compounds. Amoxicillin was used as the reference compound to compare the binding energies and interaction profiles. In the wild-type PBP2x, Amoxicillin exhibited a binding energy of -7.73 kcal/mol. For the STMK motif mutations, T338A, T338G, and T338P, the binding energies of Amoxicillin were -7.79 kcal/mol, -6.86 kcal/mol, and -7.76 kcal/mol, respectively. In case of KSG mutations K547G and K547T, the binding energies were observed to be -7.77 kcal/mol, -7.28 kcal/mol, respectively, when docked with Amoxicillin. Hydrogen bond analysis revealed that in the wild-type PBP2x-Amoxicillin, four hydrogen bonds were formed with the residues SER337, THR526, LYS340, and SER395. In the K547G mutant, Amoxicillin formed five hydrogen bonds involving THR550, GLY567, SER548, SER395, and ASN397, although one unfavourable bond was also observed. Similarly, in the K547T variant, Amoxicillin formed two hydrogen bonds with GLY597 and THR550. For STMK site mutations, Amoxicillin formed two hydrogen bonds with SER337 and TRP374 in the T338A mutant, including an unfavourable interaction. In the T338G mutant, four hydrogen bonds were observed with THR550, SER337, GLU378 and TRP374. The T338P mutation resulted in three hydrogen bonds with SER337, THR526 and TRP374, along with one unfavourable bond.

All five phytocompounds were then docked with the five mutated forms of PBP2x. The docking scores range from -8.22 to -2.28 kcal/mol; the values are given provided within the Supplementary Table S5. Among these, Glucozaluzanin C (Compound ID: 1063) demonstrated the lowest binding energy of -8.22 kcal/mol, indicating the strongest interaction with PBP2x. The docking results of Glucozaluzanin C revealed a total of five hydrogen bonds in the wild-type complex. Specifically, in the T338A mutant, Glucozaluzanin C formed four hydrogen bonds, primarily with SER337 and THR550. In the T338G mutant, it formed seven hydrogen bonds involving SER337, THR550, and GLU978. For the T338P mutant, Glucozaluzanin C formed six hydrogen bonds with SER395, LYS340, THR550 and SER337. In the K547G mutant, Glucozaluzanin C formed four hydrogen bonds that are THR526, GLY549, THR550, and SER571. Similarly, the K547T mutant formed 2 hydrogen bonds with THR526 and SER571. The comparative interaction profiles of Amoxicillin and Glucozaluzanin C with wild-type and mutated PBP2x proteins are presented in Supplementary Figs. S8 and S9, while the corresponding docking scores and interacting residues are summarised in Supplementary Table S6. The docking analysis demonstrated that Glucozaluzanin C shared several key binding residues with Amoxicillin, especially in the mutated binding pockets, suggesting its potential to retain binding affinity even in the presence of active site mutations. Glucozaluzanin C was found to effectively bind with the active site of mutated PBP2x and showed favourable binding energies and interactions. Based on the docking scores, interaction patterns and DFT analysis, Glucozaluzanin C was selected for further evaluation through molecular dynamic simulations.

Dynamics simulations analysis

MD simulations were performed for 100 ns to evaluate the molecular behaviour and dynamics of the top docked compound, Glucozaluzanin C, in complex with both the wild-type and mutants of PBP2x. The resulting trajectories were analysed using multiple MD parameters. The analysis provided a comprehensive estimation of the structural stability and conformational fluctuations of the complexes. The RMSD values, which indicate the structural deviation of the complex over time, showed that all systems stabilised, with RMSD values ranging from 0.21 to 0.26 nm, indicating high structural stability. RMSD values less than 0.3 nm are considered acceptable for stable simulations. The average RMSD for each complex are provided in the table, and motif-wise fluctuations. MD simulations were also performed for the apo forms of the wild-type and mutants PBP2x proteins to evaluate baseline structure dynamics. The simulation results, including RMSD, RMSF, hydrogen bond and SASA are provided in the Supplementary Fig. S10 for comparative reference. Further, RMSF analysis over the 100 ns simulation provided insights into residue-level flexibility. The wild type showed moderate fluctuations, particularly in loop regions around residue numbers ~ 300 and ~ 500 . Mutants generally displayed reduced fluctuations at catalytic motifs, indicating localised rigidity that could favour ligand binding. RMSF values for PBP2x in complex with Glucozaluzanin C ranged from 0.11 to 0.14 nm. The RMSF profile between residues ~ 92 –189 corresponds to an unresolved segment in the crystal structure. As confirmed by sequence alignment (Supplementary Fig. S11), this region lies outside the penicillin binding transpeptidase domain (residues 289–609) and does not influence the functional dynamics of the protein. Prior studies have shown that simulations of partially resolved protein remain valid when missing loops are distant from the active domain^{65,66}. Hydrogen bond analysis showed that Glucozaluzanin C formed up to six hydrogen bonds with the wild type, several of which remained stable. In the K547G mutant, five hydrogen bonds were observed, with three consistent throughout the simulation. In K547T, six bonds formed, with one bond stable throughout the simulation, and more bonds forming towards the end. In the T338A mutant, the complex formed seven hydrogen bonds, with two remaining stable. T338G showed six hydrogen bonds with three stable ones, and T338P formed seven hydrogen bonds, with three being constant.

The radius of gyration (RoG) was used to evaluate structural efficiency during the simulations, where the lower RoG values specify more stable and compact protein structures. Glucozaluzanin C-PBP2x complexes showed minimal fluctuation in RoG values, ranging from 2.42 to 2.44 nm across all systems, SASA analysis also revealed low variation in the solvent-accessible surface area, suggesting that the presence of Glucozaluzanin C helped maintain protein folding and surface stability throughout the 100ns simulation. Interaction energy calculated using short-range Lennard–Jones interactions (LJIE) further supported strong ligand binding. The wild-type complex showed an interaction energy of -112.14 KJ/mol with a 2.4 error rate. For mutants K547G and K547T, interaction energies were -127.29 KJ/mol and -120.37 KJ/mol, with error rates of 6.70 and 5.60, respectively. T338A, T338G, and T338P showed interaction energies of -101.34 , -105.76 and -141.68 KJ/mol, with corresponding errors of 3.50, 2.70 and 1.70 and the results are depicted in Fig. 3a and l. These values indicate effective binding in all systems (Table 2). PCA was conducted to study overall structural motion and conformational dynamics (Supplementary Fig. S12). FEL based on the first two principal components (PC1 and PC2) were generated. The wild-type complex revealed a deep energy minimum, indicating conformational

stability. Mutants K547G and K547T showed low energy basins with retained structural stability. The mutants T338A, T338G, and T338P showed well-formed energy wells, supporting functionally relevant conformational changes while maintaining favourable interactions with Glucozaluzanin C. These results indicate that the PBP2x-Glucozaluzanin C complexes exhibit consistent, stable dynamics and binding across wild-type and mutant systems, particularly within the transpeptidase domain (Supplementary Fig. S13).

MM-GBSA analysis

The binding free energy (ΔG) for the best protein-ligand complexes was evaluated using the gmX MMPBSA. Additionally, the gmX MMPBSA ana utility was used to perform component analysis and generate visual

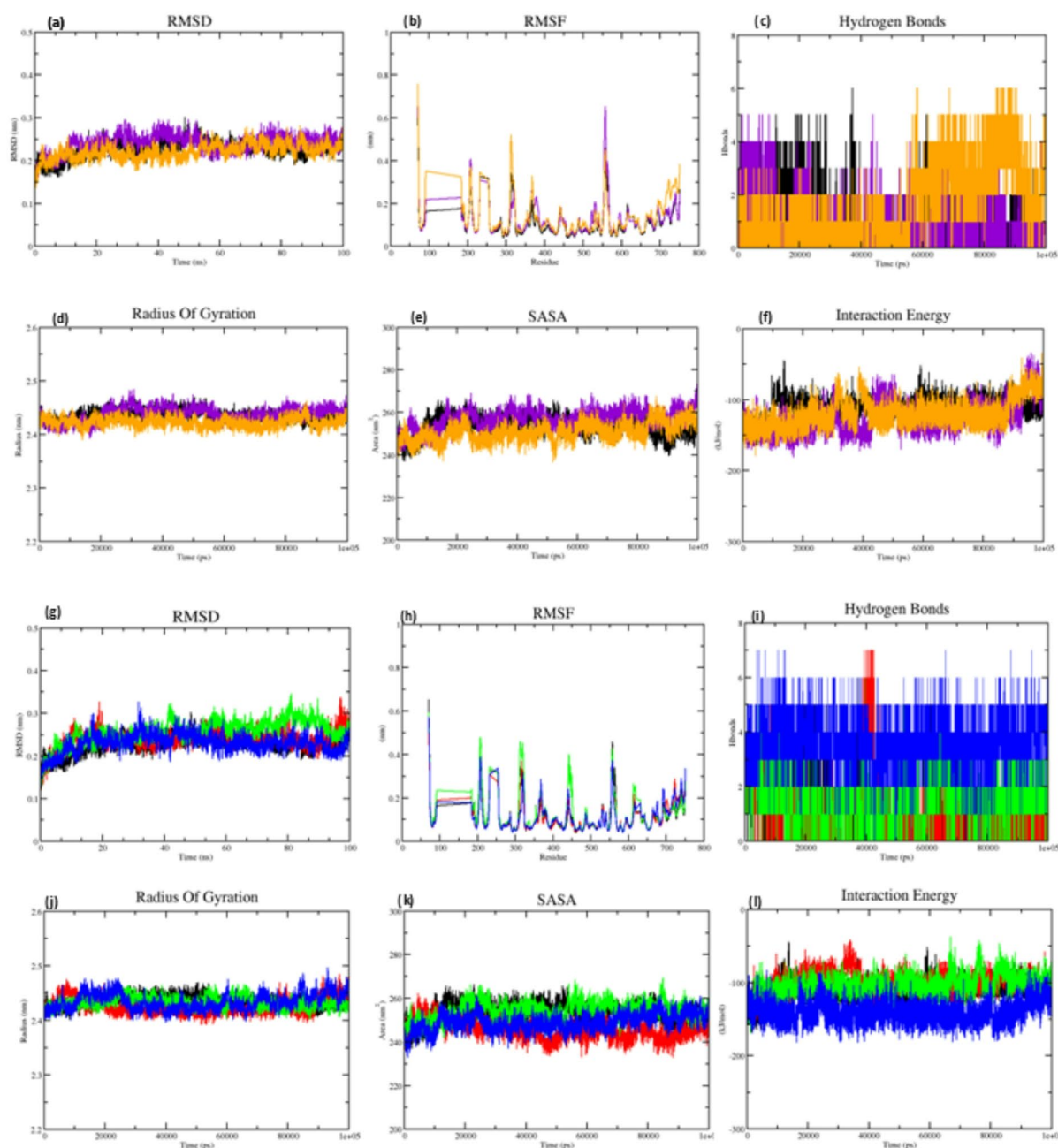


Fig. 3. Molecular dynamics simulations for KSG motif (wild-type complex (black), mutant K547G complex (violet) and mutant K547T complex (orange)) and STMK motif (wild-type complex (black), mutant T338A complex (red), mutant T338G complex (green), and mutant T338P complex (blue)) (a) RMSD of STMK motif (b) RMSF of STMK motif (c) Number Hydrogen bond of STMK motif (d) Radius of Gyration of STMK motif (e) SASA of STMK motif (f) Interaction Energy of STMK motif (g) RMSD of KSG motif (h) RMSF of KSG motif (i) Number Hydrogen bond of KSG motif (j) Radius of Gyration of KSG motif (k) SASA of KSG motif (l) Interaction Energy of KSG motif of KSG motif (c) SASA of KSG motif (e) Interaction Energy of KSG motif.

Molecular dynamics simulation parameters								MM/GBSA energy profile						
Complexes	Avg. RMSD (nm)	Avg. RMSF (nm)	Avg. ROG (nm)	Avg. ASA (nm ²)	Avg. IE (kJ/mol)	Avg. Rate	H-Bond	ΔVD -WAALS	ΔEEL	ΔEGB	ΔESURF	ΔGGAS	ΔGSOLV	ΔTotal
WT-1063	0.23	0.11	2.44	253.49	− 112.14	2.40	6	− 23.64	− 15.94	31.31	− 3.56	− 39.58	27.74	− 11.84
MT-K547G-1063	0.24	0.12	2.44	256.85	− 127.29	6.70	5	− 27.93	− 25.04	41.97	− 4.21	− 52.97	37.76	− 15.2
MT-K547T-1063	0.22	0.14	2.42	250.89	− 120.37	5.60	6	− 28.95	− 22.27	40.32	− 4.56	− 51.23	35.76	− 15.46
MT-T338A-1063	0.21	0.11	2.43	245.54	− 101.34	3.50	7	− 28.37	− 6.22	23.01	− 3.91	− 34.59	19.1	− 15.49
MT-T338G-1063	0.26	0.13	2.43	254.06	− 105.76	2.70	6	− 34.36	− 14.12	32.97	− 5	− 48.48	27.96	− 20.51
MT-T338P-1063	0.23	0.11	2.43	248.93	− 141.68	1.70	7	− 32.55	− 20.9	37.98	− 5	− 53.45	32.98	− 20.48

Table 2. Molecular dynamics simulation parameters and MM/GBSA energy profile of Glucozaluzanin C complexes with wild-type and mutant PBP2x.

representations in the form of graphs. These graphs depict the overall binding energy components, including gas-phase energy (ΔG gas), solvation energy (ΔG solvation) and total binding free energy (ΔG total), as well as specific individuals contributing such as van der Waals (ΔG EvdW) and electrostatic (ΔG Eelec) interactions. (Supplementary Fig. S14) represents comparative values for all six protein-ligand complexes evaluated. The wild-type PBP2x—Glucozaluzanin C complex exhibited a ΔG total of - 11.84 kcal/mol. Among the three mutants in the STMK motif (T338A, T338G, and T338P), the calculated ΔG total values were - 15.20 kcal/mol, - 15.46 kcal/mol and - 15. 49 kcal/mol respectively. For the KSG mutants (K547G and K547T), stronger binding affinities were observed with ΔG total values of - 20.51 kcal/mol and - 20.48 kcal/mol, respectively (Table 2) These results suggest that Glucozaluzanin C forms favourable and stable complexes with both wild-type and mutant forms of PBP2x. The improved binding affinities observed in the mutant complexes can be assigned to enhance van der Waals interactions, favourable gas-phase contributions, and supported electrostatic interactions. These findings confirm that Glucozaluzanin C exhibits acceptable ΔG values and forms stable complexes with the mutants of PBP2x. To further understand the residue-level contributions to binding, per-residue MM/GBSA energy decomposition analysis was performed. Notably, residues such as SER395, LYS496 and THR550 contributed significantly to the overall binding energy across all complexes. The residues appear to be key stabilizing elements within the binding pockets (Supplementary Fig. S15).

Discussion

PBP2x plays a crucial role in the development of the β -lactam resistance in *S. pneumoniae*, primarily due to its essential function in catalysing the transpeptidation reaction during bacterial cell wall biosynthesis. Mutations within the conserved motifs of PBP2x weaken its ability to bind β -lactam antibiotics, thereby reducing susceptibility and contributing to resistance. Specifically, these alterations can also affect the proteins' enzymatic function, potentially leading to defects in the cell wall synthesis¹⁰. Among the widely investigated mutations, at position T338, located adjacent to the active site S337 in the STMK catalytic motif, is frequently substituted with T338A/G/P. These mutations have been experimentally validated in vitro and are known to influence β -lactam binding⁶⁷. Similarly, mutations in the KSG catalytic motif, such as the substitution of K547 with K547G/T, occur at the end of the β 3 loop and are associated with increased acylation efficiency towards β -lactam^{7,12}. These changes, being situated in conserved active sites, can substantially alter the protein's structural and functional integrity. Recent regional studies reported that PBP2x mutations, particularly T338 and K547 substitutions, are strongly associated with respiratory and invasive infections such as pneumonia, meningitis and septicaemia. Serotypes 19A, 6b and 23F, which carry these mutations, are the most frequently detected in clinical cases across different countries such as China, India, Japan, South Korea, Vietnam, and parts of Europe, where antimicrobial resistance is a growing public health challenge. The widespread presence of these resistant variants highlights the critical need for new inhibitors targeting the structurally altered PBP2x⁶⁸.

In this study, we specifically analysed mutations within the STMK and KSG motifs, particularly T338A/G/P, K547G/T. All selected mutations have been reported in the literature as clinically relevant and tested. Computational evaluation indicated that each of these mutations had a deleterious effect on PBP2x functionality. These findings highlight the urgent need for novel therapeutic agents targeting β -lactam resistant *S. pneumoniae* strains. The increasing prevalence of multidrug-resistant organisms is largely driven by the widespread use of antibiotics, which accelerates the selection of resistance factors⁶⁹. Natural compounds, especially those derived from plants, offer a promising alternative for antimicrobial drug discovery due to their broad-spectrum biological activities. These include antibacterial, antioxidant, anti-inflammatory, anticancer, immunomodulatory, and neuroprotective effects⁷⁰. Several phytochemicals have demonstrated efficacy against *S. pneumoniae*, particularly in conjunction with or as alternatives to β -lactam antibiotics^{71,72}. Phenolic compounds are of particular interest because of their ability to bind to bacterial virulence factors such as PBPs, inhibit biofilm formation, and disrupt bacterial membranes. These effects have been documented in Gram-positive and Gram-negative bacteria^{73,74}. In this study, we employed ML models to screen a library of plant-based compounds against mutant variants of PBP2x. The application of ML improved in the identification of phytochemicals with potential inhibitory activity against all five resistance-associated PBP2x mutations.

Previous studies have demonstrated the effectiveness of ML in predicting active compounds against protein targets associated with drug resistance^{20,75}. The use of ROC curves and the AUC-ROC provided robust metrics for model performance evaluation, distinguishing active from inactive compounds⁷⁶. Genomic surveillance

tools, as shown by Bilal and team, show how unsupervised ML algorithms can be used to study evolutionary patterns and predict susceptibility based on PBP sequence variation⁷⁷. In our study, ML algorithms were employed to filter active compounds based on their predicted activity profiles, further evaluated through ADMET profiling DFT analysis. Among the top compounds, Glucozaluzanin C, a guaiane-type sesquiterpene lactone isolated from *Elephantopus scaber*, showed the most favourable binding energy and was selected for further molecular docking and simulation studies. DFT analysis for all five lead compounds revealed low HOMO-LUMO energy gaps, indicative of high chemical reactivity and potential biological activity. In particular, the DFT-derived electronic properties of the ligands offer detailed insights into their biological activity and binding behaviour. Compound Glucozaluzanin C, which exhibited a high HOMO-LUMO energy gap (5.69 eV), suggests high polarizability and stability. It also demonstrated strong binding in docking (-8.22 kcal/mol) and favourable MM/GBSA binding free energy (-11.84 kcal/mol). The high dipole moment (7.3D) further specifies enhanced complementary electrostatic interaction with charged residues in the binding pockets. ESP analysis showed electron-rich zone near hydroxyl and carbonyl groups, aligning well with the hydrogen-bond donor/acceptor regions in PBP2x. Per-residue MM/GBSA decomposition analysis confirmed consistent contribution to binding stabilization. These results support a detailed understanding of how electronic descriptors explain PBP2x – Glucozaluzanin C complex interaction. MD simulations further confirmed the structural stability of PBP2x – Glucozaluzanin C complex, as indicated by the consistent RMSD, RMSF, RoG and SASA over the period^{78,79}. To better understand the global motion and dynamic behaviour of the protein-ligand complex, PCA was performed on the trajectory. The first few eigenvectors accounted for the majority of the collective motions, confirming that Glucozaluzanin C binding results in a stable and restricted conformational landscape compared to the unbound state. Furthermore, MM/GBSA free energy calculations were employed to estimate the binding affinity, indicating highly favourable interactions, particularly in T338P and K547G variants, suggesting strong binding affinity and stability of the complex. Glucozaluzanin C has previously exhibited antimicrobial effects against *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Bacillus subtilis*, and *Aspergillus niger*^{80–83}. Although direct evidence of its efficacy against *S. pneumoniae* remains limited, its traditional use and broad-spectrum antimicrobial properties suggest the need for further investigation⁸⁴. ADMET evaluation confirmed Glucozaluzanin C's favourable pharmacokinetic profile, including good absorption, non-toxic behaviour and compatibility as a P-glycoprotein substrate. While these computational findings are encouraging, experimental validation is still needed. Glucozaluzanin C has not yet been tested in vitro and in vivo against *S. pneumoniae*, and its inhibitory activity remains to be confirmed through biological studies. Future work will focus on validating the efficacy of the compound using in vivo experimental models.

Overall, this study demonstrated that Glucozaluzanin C exhibits strong potential as an inhibitor of β -Lactam resistant PBP2x in *S. pneumoniae*. The integration of ML-based screening, DFT electronic characterisation, MM/GBSA energetic evaluation, PCA-based motion analysis, and MD simulation presents a comprehensive and effective strategy for natural product-based drug discovery. Compared to traditional QSAR approaches, ML models offer greater accuracy and efficiency with smaller datasets^{85,86}. This emphasizes the values of ML-driven pipelines for early-stage identification of phytochemical leads in the context of antibiotic resistance. DFT analysis provided additional insights into the electronic properties and reactivity of Glucozaluzanin C, with moderate HOMO-LUMO energy gaps indicating favourable electron transfer characteristics and potential biological activity⁸⁷. The absence of imaginary vibrational frequencies confirmed that the optimised geometry represents a true energy minimum, supporting the structural stability of the compound⁸⁸. Electrostatic potential surface mapping further revealed key regions of nucleophilic and electrophilic potential, which are relevant to interactions with the PBP2x active site⁷⁸. Geometry optimisation ensured a reliable molecular conformation for downstream analysis. Finally, molecular docking simulations targeting key PBP2x resistance mutations (T338A, T338G, T338P, K547G, K547T) supported that Glucozaluzanin C forms stable hydrogen bonds as a competitive inhibitor^{89,90}.

Conclusion

The findings of this study revealed the important mutations (T338A/G/P and K547G/T) in the PBP2x protein of *S. pneumoniae*, which are known to reduce the effectiveness of β -lactam antibiotics. Using a combination of machine learning, DFT analysis, molecular docking, and dynamic simulations, we screened and evaluated plant-based compounds for their ability to bind to the resistant protein PBP2x. Among the compounds tested, Glucozaluzanin C showed the most promising results. It was able to bind effectively to all five PBP2x mutants, showed stable behaviours during simulations, and had favourable electronic and pharmacokinetic properties. Its known antimicrobial activity and strong interaction with PBP2x mutants make it a good candidate for further study. The computational approach used in this study provides a useful method for identifying natural compounds that could help in the antibiotic-resistant organisms. These results may support the future development of Glucozaluzanin C and related molecules as therapeutic agents against β -lactam resistant *S. pneumoniae*.

Data availability

The protein structure analysed for this study was retrieved from the Protein Data Bank (PDB) <https://www.rcsb.org/>, and the phytocompounds were obtained from the IMPPAT Database <https://cb.imsc.res.in/imppat/>.

Received: 2 June 2025; Accepted: 13 October 2025

Published online: 07 November 2025

References

1. Evans, W. & Hansman, D. Tetracycline-resistant pneumococcus. *Lancet* **281**, 451 (1963).

2. Zahari, N. I. N. et al. A review of the resistance mechanisms for β -Lactams, macrolides and fluoroquinolones among *Streptococcus pneumoniae*. *Med. (B Aires)*. **59**, 1927 (2023).
3. Varghese, R. et al. Analysis of amino acid sequences of penicillin-Binding proteins 1a, 2b, and 2x in invasive *Streptococcus pneumoniae* nonsusceptible to penicillin isolated from children in India. *Microb. Drug Resist.* **27**, 311–319 (2021).
4. Pernot, L. et al. A PBP2x from a clinical isolate of *Streptococcus pneumoniae* exhibits an alternative mechanism for reduction of susceptibility to β -Lactam antibiotics. *J. Biol. Chem.* **279**, 16463–16470 (2004).
5. Hakenbeck, R., Brückner, R., Denapate, D. & Maurer, P. Molecular mechanisms of β -Lactam resistance in *Streptococcus pneumoniae*. *Future Microbiol.* **7**, 395–410 (2012).
6. Chochua, S. et al. Invasive group A streptococcal penicillin binding protein 2x variants associated with reduced susceptibility to β -Lactam antibiotics in the United States, 2015–2021. *Antimicrob. Agents Chemother.* **66**, e00802–e00822 (2022).
7. Jensen, A., Valdósson, O., Frimodt-Møller, N., Hollingshead, S. & Kilian, M. Commensal streptococci serve as a reservoir for β -Lactam resistance genes in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **59**, 3529–3540 (2015).
8. Aribisala, J. O., S'thebe, N. W. & Sabiu, S. In silico exploration of phenolics as modulators of penicillin binding protein (PBP) 2x of *Streptococcus pneumoniae*. *Sci. Rep.* **14**, 8788 (2024).
9. Dessen, A., Mouz, N., Gordon, E., Hopkins, J. & Dideberg, O. Crystal structure of PBP2x from a highly penicillin-resistant *Streptococcus pneumoniae* clinical isolate. *J. Biol. Chem.* **276**, 45106–45112 (2001).
10. Mouz, N. et al. Mutations in the active site of penicillin-binding protein PBP2x from *Streptococcus pneumoniae*. *J. Biol. Chem.* **274**, 19175–19180 (1999).
11. Panickar, A., Manoharan, A. & Ramaiah, S. Single nucleotide polymorphisms and penicillin non-susceptibility among invasive *Streptococcus pneumoniae* from Vietnam and India: Insights from a comparative genomics study. *J. Glob. Antimicrob. Resist.* **43**, 120–129 (2025).
12. Zhou, M. et al. Molecular characterization of penicillin-binding protein 2x, 2b and 1a of *Streptococcus pneumoniae* causing invasive pneumococcal diseases in China: A multicenter study. *Front. Microbiol.* **13**, 838790 (2022).
13. Basu, S. et al. Non-steroidal anti-inflammatory drugs ketorolac and etodolac can augment the treatment against pneumococcal meningitis by targeting penicillin-binding proteins. *Microb. Pathog.* **170**, 105694 (2022).
14. Zhang, C. et al. Systematic analysis of supervised machine learning as an effective approach to predicate β -lactam resistance phenotype in *Streptococcus pneumoniae*. *Brief. Bioinform.* **21**, 1347–1355 (2020).
15. Ghosh, S., Basu, S., Anbarasu, A. & Ramaiah, S. A. Comprehensive review of antimicrobial agents against clinically important bacterial pathogens: Prospects for phytochemicals. *Phyther Res.* **39**, 138–161 (2025).
16. Kaya, D. E., Ülgen, E., Kocagöz, A. S. & Sezerman, O. U. A comparison of various feature extraction and machine learning methods for antimicrobial resistance prediction in *Streptococcus pneumoniae*. *Front. Antibiot.* **2**, 1126468 (2023).
17. Gurbych, O., Druchok, M., Yarish, D. & Garkot, S. High throughput screening with machine learning (2020).
18. Gibson, P. S., Bexkens, E., Zuber, S., Cowley, L. A. & Veening, J. W. The acquisition of clinically relevant amoxicillin resistance in *Streptococcus pneumoniae* requires ordered horizontal gene transfer of four loci. *PLOS Pathog.* **18**, e1010727 (2022).
19. Grebe, T. & Hakenbeck, R. Penicillin-binding proteins 2b and 2x of *Streptococcus pneumoniae* are primary resistance determinants for different classes of beta-lactam antibiotics. *Antimicrob. Agents Chemother.* **40**, 829–834 (1996).
20. Mathpal, S., Joshi, T., Priyavada, P., Ramaiah, S. & Anbarasu, A. Machine learning and cheminformatics-based identification of lichen-derived compounds targeting mutant PBP4R200L in *Staphylococcus aureus*. *Mol. Divers.* **29**, 3345–3370. <https://doi.org/10.1007/s11030-025-11125-6> (2025).
21. Choi, Y. & Chan, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
22. Capriotti, E., Altman, R. B. & Bromberg, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genom.* **14**(Suppl 3), S2 (2013).
23. Ashkenazy, H. et al. ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
24. Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Sci.* **30**, 60–69 (2021).
25. Mohanraj, K. et al. IMPPAT: A curated database of Indian medicinal plants, phytochemistry and therapeutics. *Sci. Rep.* **8**, 4329 (2018).
26. Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **32**, 1466–1474 (2011).
27. Guyon, I. A. E. An introduction of variable and feature selection. *Journal Mach. Learn. Res.* **3**, 157–1182 (2003).
28. Hall, M. et al. The WEKA data mining software. *ACM SIGKDD Explor. Newsl.* **11**, 10–18 (2009).
29. Frank, E., Hall, M. A. & Witten, I. H. The WEKA workbench. *Data Min.* <https://doi.org/10.1016/b978-0-12-804291-5.00024-6> (2017).
30. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **7**, 91 (2006).
31. Chen, R., Liu, X., Jin, S., Lin, J. & Liu, J. Machine learning for drug-target interaction prediction. *Molecules* **23**, 2208 (2018).
32. Lavecchia, A. Machine-learning approaches in drug discovery: Methods and applications. *Drug Discov. Today* **20**, 318–331 (2015).
33. Roy, A. & Anbarasu, A. Unveiling Berberine analogues as potential inhibitors of *Escherichia coli* FtsZ through machine learning molecular docking and molecular dynamics approach. *Sci. Rep.* **15**, 14668 (2025).
34. van de Waterbeemd, H. & Gifford, E. ADMET in silico modelling: Towards prediction paradise? *Nat. Rev. Drug Discov.* **2**, 192–204 (2003).
35. Fu, L. et al. ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Res.* **52**, W422–W431 (2024).
36. Cheng, F. et al. AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* **52**, 3099–3105 (2012).
37. Banerjee, P., Kemmler, E., Dunkel, M. & Preissner, R. ProTox 3.0: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* **52**, W513–W520 (2024).
38. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. PkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **58**, 4066–4072 (2015).
39. Geerlings, P., Fias, S., Boisdenghien, Z. & De Proft, F. Conceptual DFT: Chemistry from the linear response function. *Chem. Soc. Rev.* **43**, 4989 (2014).
40. Chai, J. D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **128**, 084106 (2008).
41. Chattaraj, P. K. & Giri, S. Electrophilicity index within a conceptual DFT framework. *Annu. Rep. Sect. 'C' (Phys. Chem.)* **105**, 13 (2009).
42. Naha, A. et al. Network metrics, structural dynamics and density functional theory calculations identified a novel ursodeoxycholic acid derivative against therapeutic target parkin for Parkinson's disease. *Comput. Struct. Biotechnol. J.* **20**, 4271–4287 (2022).
43. Morris, G. M. et al. AutoDock4 and AutoDockTools4: Automated Docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
44. Santos-Martins, D., Forli, S., Ramos, M. J. & Olson, A. J. AutoDock4_{v4}: an improved AutoDock force field for small-molecule docking to zinc metalloproteins. *J. Chem. Inf. Model.* **54**, 2371–2379 (2014).

45. Huey, R., Morris, G. M., Olson, A. J. & Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28**, 1145–1152 (2007).
46. Mathpal, S. et al. In Silico screening of chalcone derivatives as promising EGFR-TK inhibitors for the clinical treatment of cancer. *3 Biotech.* **14**, 18 (2024).
47. Meng, X. Y., Zhang, H. X., Mezei, M. & Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided-Drug Des.* **7**, 146–157 (2011).
48. Joshi, T. et al. Repurposing of FDA approved drugs against *Salmonella enteric* serovar Typhi by targeting dihydrofolate reductase: an in silico study. *J. Biomol. Struct. Dyn.* **40**, 3731–3744 (2022).
49. Flanders, P. L. et al. Combined structural analysis and molecular dynamics reveal penicillin-binding protein Inhibition mode with β -Lactones. *ACS Chem. Biol.* **17**, 3110–3120 (2022).
50. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* **1**, 19–25 (2015).
51. Savita, B. K. et al. Characterization of recombinant pumpkin 2S albumin and mutation studies to unravel potential DNA/RNA binding site. *Biochem. Biophys. Res. Commun.* **580**, 28–34 (2021).
52. da Silva, T. U., Pougy, K., de Albuquerque, C., da Silva Lima, M. G., de Machado, S. & C. H. & Development of parameters compatible with the CHARMM36 force field for $[\text{Fe}_4\text{S}_4]^{2+}$ clusters and molecular dynamics simulations of adenosine-5'-phosphosulfate reductase in GROMACS 2019. *J. Biomol. Struct. Dyn.* **40**, 3481–3491 (2022).
53. Mathpal, S. et al. Screening of potential bio-molecules from *Moringa olifera* against SARS-CoV-2 main protease using computational approaches. *J. Biomol. Struct. Dyn.* **40**, 9885–9896 (2022).
54. Pronk, S. et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
55. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101. <https://doi.org/10.1063/1.2408420> (2007).
56. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
57. Panchalingam, S., Jayaraman, M., Jeyaraman, J. & Kasivelu, G. Harnessing marine natural products to inhibit PAD4 triple mutant: A structure-based virtual screening approach for rheumatoid arthritis therapy. *Arch. Biochem. Biophys.* **761**, 110164 (2024).
58. George, E. A. et al. Pharmacokinetics Screening, molecular Docking, and dynamics simulations revealed novel antimicrobial peptide NKLf2 mutants as potent inhibitors of *Mycobacterium tuberculosis*. *Probiotics Antimicrob. Proteins*. <https://doi.org/10.1007/s12602-025-10608-0> (2025).
59. Jayaraman, M., Gosu, V., Kumar, R. & Jeyaraman, J. Computational insights into potential marine natural products as selective inhibitors of *Mycobacterium tuberculosis* inhA: A structure-based virtual screening study. *Comput. Biol. Chem.* **108**, 107991 (2024).
60. Jayaraman, M., Rajendra, S. K. & Ramadas, K. Structural insight into conformational dynamics of non-active site mutations in kasa: A *Mycobacterium tuberculosis* target protein. *Gene* **720**, 144082 (2019).
61. Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A. & Moreno, E. gmx_MMPBSA: A new tool to perform End-State free energy calculations with GROMACS. *J. Chem. Theory Comput.* **17**, 6281–6291 (2021).
62. Joshi, T. et al. Identifying novel therapeutics for the resistant mutant F533L in PBP3 of *Pseudomonas aeruginosa* using ML techniques. *ACS Omega.* **9**, 28046–28060 (2024).
63. Singh, V. et al. Drug-Repurposing approach to combat *Staphylococcus aureus*: biomolecular and binding interaction study. *ACS Omega.* **7**, 38448–38458 (2022).
64. Dalal, V., Golemi-Kotra, D. & Kumar, P. Quantum mechanics/molecular mechanics studies on the catalytic mechanism of a novel esterase (FmtA) of *Staphylococcus aureus*. *J. Chem. Inf. Model.* **62**, 2409–2420 (2022).
65. Skliros, A., Jernigan, R. L. & Kloczkowski, A. Models to approximate the motions of protein loops. *J. Chem. Theory Comput.* **6**, 3249–3258 (2010).
66. Powell, A. J., Tomberg, J., Deacon, A. M., Nicholas, R. A. & Davies, C. Crystal structures of penicillin-binding protein 2 from penicillin-susceptible and -resistant strains of neisseria gonorrhoeae reveal an unexpectedly subtle mechanism for antibiotic resistance. *J. Biol. Chem.* **284**, 1202–1212 (2009).
67. Zerfaß, I., Hakenbeck, R. & Denapaite, D. An important site in PBP2x of penicillin-resistant clinical isolates of *Streptococcus pneumoniae*: mutational analysis of Thr338. *Antimicrob. Agents Chemother.* **53**, 1107–1115 (2009).
68. Kim, S. H. et al. Changing trends in antimicrobial resistance and serotypes of *Streptococcus pneumoniae* isolates in Asian countries: An Asian network for surveillance of resistant pathogens (ANSORP) study. *Antimicrob. Agents Chemother.* **56**, 1418–1426 (2012).
69. Varghese, R. et al. Emergence of meropenem resistance among cefotaxime Non-susceptible *Streptococcus pneumoniae*: Evidence and challenges. *Front. Microbiol.* **12**, 810414 (2022).
70. Savickiene, N. & Raudone, L. Trends in plants phytochemistry and bioactivity analysis. *Plants* **13**, 3173 (2024).
71. Ghosh, S. et al. Computational advancements to facilitate therapeutic application of phytochemicals: Where do we stand? *Discov Appl. Sci.* **7**, 491 (2025).
72. Basu, S., Veeraraghavan, B. & Anbarasu, A. Anti-bacterial compounds from Indian curry-leaf tree *Murraya koenigii* have potential to inhibit carbapenem-resistant *Streptococcus pneumoniae*. *Clin. Epidemiol. Glob. Health* **28**, 101511 (2024).
73. Alhadrami, H. A. et al. Flavonoids as potential anti-MRSA agents through modulation of PBP2a: A computational and experimental study. *Antibiotics* **9**, 562 (2020).
74. Araya-Cloutier, C., Vincken, J. P., van Ederen, R., den Besten, H. M. W. & Gruppen, H. Rapid membrane permeabilization of *Listeria monocytogenes* and *Escherichia coli* induced by antibacterial prenylated phenolic compounds from legumes. *Food Chem.* **240**, 147–155 (2018).
75. Gawalska, A. et al. Application of automated machine learning in the identification of multi-target-directed ligands blocking PDE4B, PDE8A, and TRPA1 with potential use in the treatment of asthma and COPD. *Mol. Inform.* **42**, 2200214 (2023).
76. Rahman, A. S. M. Z. et al. A machine learning model trained on a high-throughput antibacterial screen increases the hit rate of drug discovery. *PLOS Comput. Biol.* **18**, e1010613 (2022).
77. Bilal, H. et al. The role of artificial intelligence and machine learning in predicting and combating antimicrobial resistance. *Comput. Struct. Biotechnol. J.* **27**, 423–439 (2025).
78. Usharani, N. et al. Green synthesis and characterization of water soluble nanocarnosine: A prospective drug delivery system. *Appl. Mater. Today.* **32**, 101812 (2023).
79. Basu, S. et al. Cefiderocol susceptibility endows hope in treating carbapenem-resistant *Pseudomonas aeruginosa*: insights from in vitro and in Silico evidence. *RSC Adv.* **14**, 21328–21341 (2024).
80. Jenny, A. et al. Antibacterial activity of aerial part of extract of *Elephantopus scaber* Linn. *Bull. Pharm. Res.* **2**, 38–41 (2012).
81. Kumar, S. S., Perumal, P., Suresh, B., Antibacterial studies on leaf extract of *Elephantopus scaber* Linn. *Anc. Sci. Life* **23**, 6–8 (2004).
82. Subhash, S. & Panda, P. K. Therapeutic effects of *Elephantopus scaber* Linn.: A review. *J. Ayurveda Holist. Med.* **11**, 69–76 (2023).
83. Meng, J. C., Zhu, Q. X. & Tan, R. X. New antimicrobial mono- and sesquiterpenes from *Soroseris hookeriana* Subsp. *Erysimoides*. *Planta Med.* **66**, 541–544 (2000).
84. Yan, Q. L. et al. Sesquiterpene lactones from *Elephantopus scaber* exhibit cytotoxic effects on glioma cells by targeting GSTP1. *Bioorg. Chem.* **129**, 106183 (2022).
85. Cherkasov, A. et al. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **57**, 4977–5010 (2014).

86. D'Souza, S., Prema, K. V. & Balaji, S. Machine learning models for drug–target interactions: Current knowledge and future directions. *Drug Discov. Today*. **25**, 748–756 (2020).
87. Luo, H. J., Wang, J. Z., Deng, W. Q. & Zou, K. DFT calculations and docking study on sesquiterpene lactones: Inhibition of aromatase. *Procedia Environ. Sci.* **8**, 446–450 (2011).
88. Liu, Y. S., Yuan, W. S., Liu, Q. J., Liu, F. S. & Liu, Z. T. Structural, mechanical, electronic, vibrational properties and hydrogen bonding of a novel energetic ionic 5, 5'-dinitroamino-3, 3'-azo-oxadiazole 4, 7-diaminopyridazino [4, 5-c] Furoxan salt. *J. Mol. Model.* **30**, 321 (2024).
89. Maurer, P. et al. Penicillin-binding protein 2x of *Streptococcus pneumoniae*: Three new mutational pathways for remodelling an essential enzyme into a resistance determinant. *J. Mol. Biol.* **376**, 1403–1416 (2008).
90. Peters, K., Schweizer, I., Hakenbeck, R. & Denapate, D. New insights into beta-lactam resistance of *Streptococcus pneumoniae*: Serine protease HtrA degrades altered penicillin-binding protein 2x. *Microorganisms* **9**, 1685 (2021).

Acknowledgements

The authors would like to thank the Vellore Institute of Technology for providing the necessary facilities to carry out the work. The authors would like to thank the Indian Council of Medical Research (ICMR), New Delhi, for the research grant IRIS ID: 2021-11889; AMR/Adhoc/290/2022-ECD-II. The authors would like to thank Dr. Anand Anbarasu for his guidance in designing the research work.

Author contributions

A.P. Literature search, data curation, and writing the original draft. A.M. Study conceptualisation, Supervising, reviewing, and editing. S.R. Study conceptualisation and funding acquisition.

Funding

Open access funding provided by Vellore Institute of Technology. This work was supported by the Indian Council of Medical Research (ICMR), Government of India Agency for the research grant (IRIS ID: 2021–11889).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-24222-1>.

Correspondence and requests for materials should be addressed to S.R.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025