Received: 23 December 2025 | Revised: 12 April 2026 | Accepted: 3 June 2026 | Published online: 17 June 2026

RESEARCH ARTICLE

Medinformatics
2026, Vol. 00(00) 1–15
DOI: 10.47852/bonviewMEDIN62028879

BON VIEW PUBLISHING

Sanjeevini 3.0: An Enhanced Comprehensive Automated Web Server for Computer-Aided Drug Design

Dheeraj Kumar Chaurasia[1, 2], Aman Sharma[2, 3], Madhvi Mishra[2], Akanksha Kesharwani[2], Raushan Anjum[2], Shashank Shekhar[2], Aditya Mittal[3] and B. Jayaram[2, 4,*]

[1] *School of Interdisciplinary Research, Indian Institute of Technology Delhi, India*

[2] *Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology Delhi, India*

[3] *Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, India*

[4] *Department of Chemistry, Indian Institute of Technology Delhi, India*

Abstract: *Sanjeevini* 3.0 represents a significant advancement in the realm of structure-based lead molecule discovery. This enhanced physicochemical principles-based, machine learning-augmented comprehensive software suite/web server version comprises various modules, including RASPD+ and SEARCH-ML (for a rapid virtual screening of large libraries of small molecules against a specified target), AADS and ParDOCK+ (for active site prediction and docking), StackTox (for toxicity prediction of small molecules), and BAPPL+ (for scoring and estimating binding free energies). The pipeline addresses the complex challenges that the global pharmaceutical industry faces in translating a molecular-level understanding of human diseases into a cutting-edge technology for generating suggestions on candidate drug molecules. Each of the modules and the entire pipeline are thoroughly validated on some major life-threatening diseases, utilizing a dataset of 120 FDA-approved drugs and the corresponding 126 pharmacologically active targets. Remarkably, the entire pipeline requires only 13–15 min to predict candidate drugs for a single target protein. In nearly 90% of the cases, the pipeline successfully rediscovers known FDA-approved drugs for the target proteins. This methodology offers an efficient and streamlined approach without compromising effectiveness. *Sanjeevini* 3.0 is freely accessible at https://scfbioiitd.res.in/Sanjeevini/ and https://scfbio.iitd.ac.in/sanjeevini/.

Keywords: *Sanjeevini*, virtual screening, computer-aided drug design (CADD), machine learning, toxicity prediction

1. Introduction

Sanjeevini 3.0 represents the latest evolution of the *Sanjeevini* framework, integrating machine learning-enabled screening strategies with enhanced support for metalloprotein systems. The discovery of effective drug candidates has historically been a labor-intensive and costly process [1–3] characterized by iterative cycles of synthesis and experimental testing of large numbers of molecules [4]. To address these challenges, *in silico* approaches have emerged as powerful alternatives, enabling computational prediction and analysis of interactions between potential drug molecules and biological targets [5]. This paradigm shift has substantially accelerated lead identification by narrowing the chemical search space prior to experimental validation. Consequently, computer-aided drug design (CADD) has become an essential component of modern drug discovery pipelines, offering the potential to reduce developmental costs and time to market [6–8]. In practice, these computational tools are predominantly employed during hit identification and early hit-to-lead optimization [9], while more computationally intensive methods such as free-energy perturbation or QM/MM approaches are typically reserved for later stages of lead refinement [10].

Sanjeevini 3.0 is an advanced software suite and web server developed to tackle the increasing complexities in structure-based drug discovery. Building upon the foundations laid by its earlier versions [11], the new release incorporates upgraded algorithms, expanded small molecule databases, and validated machine learning components to enhance the efficiency and accuracy of the overall pipeline. Key improvements in this version include a more reliable treatment of metalloprotein–ligand interactions, an improved rapid screening strategy capable of evaluating millions of molecules within minutes, and a preliminary version of the toxicity predictor, addressing a long-standing challenge in CADD. In response to the growing need for effective strategies to translate

***Corresponding author:** B. Jayaram, Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology Delhi and Department of Chemistry, Indian Institute of Technology Delhi, India. Email: bjayaram@chemistry.iitd.ac.in

© The Author(s) 2026. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

molecular insights into viable therapeutic candidates [12], *Sanjeevini* 3.0 offers a comprehensive pathway for lead molecule development through its integrated modules for virtual screening, active site prediction, docking, and binding-energy estimation.

A central challenge in CADD lies in the vastness of chemical and biological space, encompassing approximately 1060 possible carbon-based molecules with molecular weights below 500 Da [13], alongside more than 240,000 experimentally resolved protein structures available in the Protein Data Bank [14]. Although this diversity offers enormous opportunities for drug discovery, the pharmaceutical industry continues to face low productivity, driven by high attrition rates, escalating costs, and aging development pipelines [15]. Additional challenges include the accurate prediction of binding affinities, the application of physics-based free-energy decomposition methods, and a reliable treatment of metal ions during docking and scoring [16]. Metal ions and cofactors play critical roles in modulating protein structure and function [17]; however, their inclusion in docking simulations and affinity predictions remains computationally demanding and methodologically complex.

Numerous software tools have been developed to address screening, docking, and scoring of small molecules [18, 19] including widely used platforms such as AutoDock [20], DOCK [21], FLOG [22], ParDOCK, and ParaDockS [23], each employing distinct strategies for ligand-receptor interaction prediction. While these tools have made significant contributions, efficient handling of metal ions, integration of machine learning techniques, and large-scale computational efficiency remain active areas of development [24]. To address these limitations, machine learning methodologies have increasingly been incorporated into CADD workflows, offering improved modeling of complex molecular interactions and enhanced scalability for large chemical libraries [25].

Building upon earlier versions of *Sanjeevini* [11], the current release integrates advanced virtual screening, docking, and scoring strategies within a unified structure-based drug discovery framework. The suite integrates diverse ligand databases, including ZINC [26], Bioactivity of Indian Medicinal Plant (BIMP) [27] (https://scfbio.iitd.ac.in/bimp/), and a curated collection of FDA-approved drugs [28], thereby supporting both de novo lead identification and drug repurposing studies [29]. The availability of chemically diverse libraries enables systematic screening across multiple targets, increasing the likelihood of identifying viable lead compounds that may be overlooked by conventional approaches. *Sanjeevini* 3.0 comprises multiple modular tools for active site prediction, high-throughput virtual screening (HTVS), docking, scoring, and post-processing analyses, with the currently available modules summarized in Table 1.

The present study documents the capabilities and performance of *Sanjeevini* 3.0, highlighting methodological advances in scoring functions, incorporation of large-scale chemical libraries, parallelized docking workflows, and explicit treatment of metal ion interactions during docking and scoring. These enhancements are achieved through state-of-the-art machine learning implementations and are supported by systematic validation of individual pipeline components. *Sanjeevini* 3.0 allows users to flexibly employ selected modules or the complete workflow, depending on specific research objectives, and is delivered through an improved, user-friendly web interface (Figure 1).

2. Materials and Methods

2.1. Overview of the *Sanjeevini* 3.0 workflow

The *Sanjeevini* 3.0 software suite comprises integrated functionalities encompassing docking, scoring, rapid screening, active site prediction, and small molecule databases, all consolidated within a unified platform. The complete workflow is outlined in Figure 2.

Table 1. The list of available modules in *Sanjeevini* 3.0 along with its description

S. no.	Name of module	Function	Link
1.	Assign atomic charge	Calculate transferrable partial atomic charge—up to 4 bonds	https://www.scfbio-iitd.res.in/software/drugdesign/charge.jsp
2.	Active site prediction	Predict active site of protein	https://scfbio-iitd.res.in/AADS/
3.	BAPPL+	Binding affinity prediction tool for non-metallo and metalloprotein–ligand complex	https://www.scfbio-iitd.res.in/bappl+/index.php
4.	Drug-likeness rules and Wiener index	Checks the drug-likeness rules compliance of ligand	https://scfbio.iitd.ac.in/sanjeevini/druglikeness.php/
5.	Molecular Volume Calculator	Calculate volume of ligand	https://scfbio.iitd.ac.in/sanjeevini/volume_calculator.php
6.	StackTox	Prediction of toxicity of small molecules	https://scfbio.iitd.ac.in/StackTox/
7.	BIMP	A database of phytochemicals from Indian medicinal plants	https://scfbio.iitd.ac.in/bimp/
8.	ParDOCK+	Non-metallo and metalloprotein–ligand docking tool	https://scfbio-iitd.res.in/pardock+/
9.	RASPD+	High-throughput virtual screening tool	https://scfbio-iitd.res.in/raspd+/
10.	SEARCH-ML	Enhanced high-throughput virtual screening tool	https://scfbio.iitd.ac.in/search-ml/

Figure 1. A snapshot of the front-end interface of the *Sanjeevini* 3.0 web server

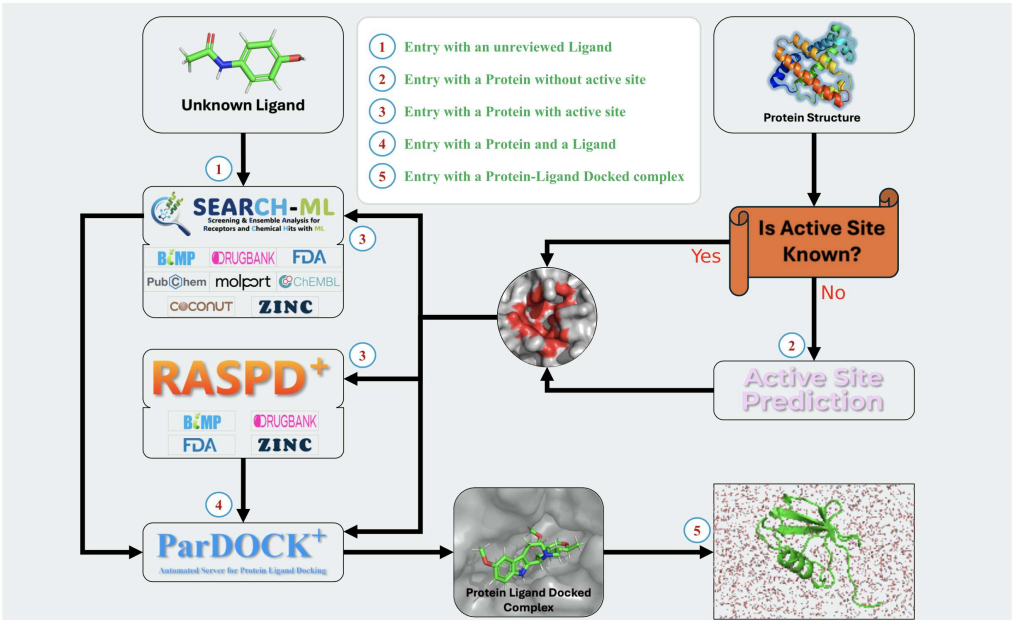

Figure 2. A schematic layout of the *Sanjeevini* 3.0 pipeline workflow

The adaptability of the pipeline and its accessibility through a range of entry points are contingent upon the extent of available information concerning the target protein and ligands. The compute cluster supporting the *Sanjeevini* platform consists of 24 nodes (768 CPU cores in total), each equipped with 96 GB RAM and running CentOS 7.4 in an OpenMPI environment. A PBS job scheduler manages job submission and resource allocation. Depending on the computational demands, different modules of *Sanjeevini* are executed in either serial or parallel mode, ensuring efficient utilization of the available high-performance computing resources. A detailed description of module integration within the workflow is provided in the following section.

2.2. Integration of pipeline modules

Sanjeevini 3.0 operates through a modular yet integrated workflow, where individual components are sequentially connected to enable efficient and biologically relevant identification of candidate drug molecules. The pipeline begins with active site prediction using the AADS module when the binding pocket is not known. This is followed by rapid screening using Screening and Ensemble Analysis for Receptors and Chemical Hits with Machine Learning (SEARCH-ML) and RASPD+, which filter large chemical libraries based on predicted interaction feasibility and binding affinity. The shortlisted compounds are then

subjected to detailed docking using the ParDOCK+ module, which generates optimal ligand conformations within the binding cavity. These docked complexes are subsequently evaluated using the Binding Affinity Prediction of Protein–Ligand Complexes (BAPPL+) scoring function to estimate binding free energies and prioritize the most promising candidates. Additionally, the StackTox module is incorporated as an early-stage filtering step to assess the toxicity of candidate molecules. Compounds predicted to be potentially toxic are removed prior to further analysis to ensure safety considerations are incorporated alongside binding affinity predictions. In cases where outputs from different modules are conflicting, such as compounds exhibiting strong binding affinity but predicted toxicity, a hierarchical decision strategy is applied. Specifically, toxicity predictions are considered a critical filtering criterion, and such compounds may be deprioritized despite favorable docking scores. This ensures that the final selection balances both efficacy and safety. The integrated design of *Sanjeevini* 3.0 allows users to either execute the complete pipeline or selectively utilize individual modules depending on the availability of input data and specific research objectives. This integration enhances the robustness of candidate selection by combining physicochemical, structural, and safety-based evaluation criteria within a unified framework.

2.3. Assignment of atomic charges using assign atomic charge tool

Sanjeevini 3.0 assigns ligand partial atomic charges using TPACM4, an empirical Generalized AMBER Force Field (GAFF) compatible charge model optimized for high-throughput docking and scoring. TPACM4 derives chemically consistent partial charges by mapping each atom in the ligand to a precompiled library of 5302 atom types generated from ~3640 small molecules. These atom types encode local chemical environments up to four bonds away and are associated with RESP-fitted HF/6-31G* charges, ensuring compatibility with biomolecular force-field parameters [30, 31]. For each input ligand, TPACM4 automatically computes bond connectivity and atom environments, identifies the closest-matching atom type in the lookup table, and assigns the corresponding partial charge. Residual charge is evenly distributed to satisfy the overall formal charge. The model reproduces RESP-level hydrogen-bond dimer energies, solvation free energies, and protein–ligand binding free energies, achieving an average error of ~1 kcal/mol across benchmark datasets. Its millisecond-level runtime per molecule enables rapid charge assignment for large chemical libraries, supporting the scale of virtual high-throughput screening performed in *Sanjeevini* 3.0.

2.4. Active site prediction (AADS)

The target protein's tertiary structure with an unidentified binding pocket is the first point of entrance. The essential prerequisite for designing a potential lead molecule to modulate protein activity is the knowledge of the binding cavity within the protein molecule. *Sanjeevini* 3.0 software suite includes the AADS module that can predict a robust binding cavity. The prediction of the binding cavity is accomplished through the application of a fuzzy scoring algorithm, which assesses the physicochemical characteristics of the functional groups that line the protein's potential binding pockets and assigns corresponding scores. It systematically incorporates multiple pertinent properties, including hydrogen-bond donor and acceptor characteristics, cavity volume, presence of ring structures, distribution of hydrophobic groups, and the occurrence of aromatic residues. A higher fuzzy score is assigned to the cavity exhibiting the highest count of hydrophobic groups, hydrogen-bond donors, and greater volume. AADS' accuracy was shown to be 100% on 620 proteins with different amino acid lengths when the top 10 projected cavities were taken into consideration.

2.5. Ligand and cavity volume calculator

The second entry point addresses proteins with known binding cavities. The molecular volume of small molecules can be determined using the "Molecular Volume Calculator" module in *Sanjeevini* 3.0. Calculating the active site and ligand volumes before docking is crucial for ensuring size compatibility between the ligand and the binding pocket, which directly influences the ligand's ability to fit and form stable interactions. A well-matched ligand can maximize key interactions, improving binding affinity and docking accuracy. Additionally, this information helps avoid steric clashes or misfitting, reduces false positives in virtual screening, and allows for better predictions of ligand orientation within the pocket. Ultimately, understanding the volumes aids in selecting ligands that are more likely to succeed [32] in the drug discovery process.

2.6. High-throughput virtual screening tool (RASPD+)

The target protein(s) can be directly subjected to HTVS using the RASPD+ module. The RASPD+ methodology represents an enhanced approach for identifying lead-like molecules, leveraging predicted binding free-energy assessments against a target protein characterized by a 3D structure and a well-defined ligand binding cavity. The methodology incorporates a comprehensive set of physicochemical descriptors for ligands, including hydrogen-bond donor and acceptor properties, partition coefficient, Weiner index, molar refractivity, and molecular weight [33]. The descriptor-based rapid screening strategy is rooted in earlier physicochemical hit-identification frameworks. Simultaneously, it assesses 14 essential physicochemical properties for the input protein. These encompass partition coefficients for both aromatic and non-aromatic residues, molar refractivity for both aromatic and non-aromatic residues, hydrogen-bond donor capacity for backbone amide groups, positively charged amino groups, neutral amino groups, heteroaromatic donors, and hydroxyl-containing groups. Furthermore, it evaluates hydrogen-bond acceptor characteristics for backbone amide, positively charged amino groups, neutral non-aromatic amino groups, and aromatic acceptors, alongside the determination of the binding pocket's volume.

In this updated version, an assortment of machine learning algorithms has been integrated, encompassing Extremely Random Forest (ERF), Random Forest (RF), k-Nearest Neighbors (KNN), Linear Support Vector Regression (SVR), Epsilon Support Vector Regression (eSVR), Linear Regression (LR), and Deep Neural Network (DNN). Notably, a novel feature for scaffold search has been introduced alongside these algorithms. Users have the flexibility to opt for either a singular method in a single run or a combination of all the available methods in a single execution. The algorithm underwent rigorous testing on a dataset comprising 493 unidentified protein–ligand complexes, following a robust training phase involving approximately 4000 non-metallo protein–ligand complexes meticulously extracted from the PDBbind refined dataset [34]. The analysis yielded notable results,

including a Pearson correlation coefficient of 0.74 and a root mean square error (RMSE) of 1.86 kcal/mol. The Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio) has developed an enhanced front-end interface for RASPD+ to facilitate user-friendly screening against a selection of pre-prepared databases, including the ZINC, DrugBank, FDA-approved, and the BIMP database, specifically customized to accommodate the input protein. RASPD+ database contains 10 million small molecules in the ZINC15 database, 8811 small molecules from the DrugBank database of Version 5.1.8, 3722 FDA-approved small molecules from DrugBank database of Version 5.1.8, and 105909 small molecules of Indian medicinal plants from the BIMP database.

Any of the four available chemical spaces can be utilized to predict the lead molecule for the target protein. RASPD+ models used within Sanjeevini 3.0 were trained on the PDBbind refined dataset (2018 release). After converting reported Kd50, Ki50, and IC 50 values to binding free energies (ΔG) and removing complexes containing metal ions within 2.1 Å of the ligand, 3925 non-covalent protein–ligand complexes were retained. A nested cross-validation strategy was used: in each of 10 replicates, 12.5% of the data (~493 complexes) served as an independent test set, while the remaining data were subjected to six-fold inner cross-validation for hyperparameter optimization. All descriptors were robustly centered and scaled using the median and interquartile range of the training fold. Final predictive performance (Pearson correlation r = 0.74, RMSE = 1.86 kcal/mol) was evaluated on an independent test set. The "Customized Dataset Screening" radio button within the RASPD+ module directs the user to a distinct input screen. Here, the users have the opportunity to construct a tailored small molecule library according to their preferences and subsequently screen their target against this ready library of choice. Users can upload either a .txt file or a .sdf file to create customized datasets of 1000 molecules at once. The RASPD+ method is publicly available, and the associated code and software can be accessed via GitHub (https://github.com/HITS-MCM/RASPDplus), and the web server is available at https://scfbio-iitd.res.in/raspd+/.

2.7. Enhanced high-throughput virtual screening and reverse virtual screening tool (SEARCH-ML)

Drawing inspiration from RASPD+, SEARCH-ML is introduced here as the first triaging layer in the *Sanjeevini* 3.0 workflow. It is designed to rapidly identify high-priority ligands from very large chemical libraries before virtual screening and docking. Unlike conventional docking-based pre-screening, SEARCH-ML employs an ensemble of machine learning models trained on experimentally validated structural data. The system was trained and validated using 11,689 protein–ligand complexes from the PDBbind.v2020 dataset, ensuring that predictions are grounded in real structural interactions. The ensemble integrates multiple state-of-the-art algorithms, including XGBoost, CatBoost, LightGBM, and RF, each contributing complementary predictive strengths [35, 36]. Model optimization was performed using Optuna-based hyperparameter tuning [37], and the final predictions are generated via a stacked ridge-regression meta-model, enabling improved generalization and stability across diverse chemical scaffolds. SEARCH-ML predicts a ligand-receptor interaction feasibility score and ranks compounds accordingly. The framework demonstrated a strong coefficient of determination (R^2 = 0.72) (Figure 3) on benchmark testing. This performance is comparable to recent deep learning-based approaches such as OnionNet-2 ($R^2 \approx 0.74$) [38], with differences arising from training scope

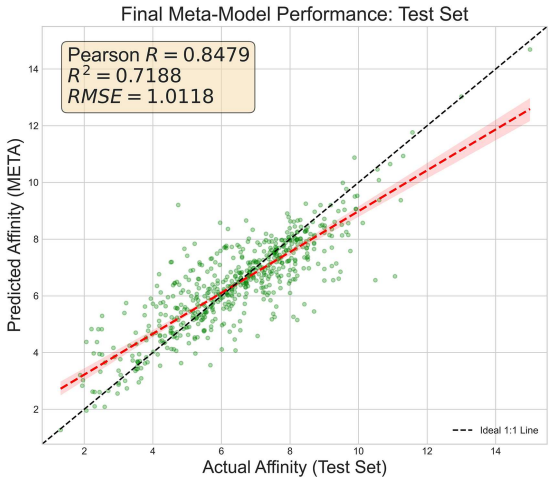

Figure 3. SEARCH-ML meta-model performance correlation plot

and design choices, as SEARCH-ML is optimized specifically for protein–ligand complexes using a newer PDBbind release and emphasizes high-throughput CPU-only screening. In addition to forward screening (protein-to-ligand), SEARCH-ML supports a reverse screening mode (ligand-to-protein), wherein users can submit a query ligand to identify and rank potential protein targets from the available receptor set. A key advantage of SEARCH-ML is its computational efficiency: the model can process 100 million compounds in under 15 minutes using CPU-only computation, significantly outperforming traditional virtual screening methods in speed and resource requirements [39]. In *Sanjeevini* 3.0, SEARCH-ML serves as an ultra-fast filtering step that reduces the search space before RASPD+ virtual screening and ParDOCK+ docking, thereby accelerating the overall pipeline while preserving prediction accuracy. The module supports screening across integrated chemical libraries including ZINC, DrugBank, FDA-approved molecules, PubChem, MolPort, ChEMBL, and BIMP and can operate even in cases where binding-site information is not explicitly known, making it broadly applicable for early-stage drug discovery workflows.

2.8. Drug-likeness filter and Wiener index calculator

To filter and screen small molecules containing the active scaffold from any of the pre-prepared libraries, users can input the SMILES representation of the scaffold either by uploading a .txt file containing a single scaffold or by directly pasting the SMILES string into the designated input field. Prior to customized dataset screening, Sanjeevini 3.0 enables evaluation of drug-likeness using multiple established physicochemical filtering criteria, including the Lipinski, Egan, Muegge, Veber, and Ghose rules. These complementary filters assess molecular properties relevant to oral bioavailability and pharmacokinetic suitability and allow systematic exclusion of compounds that do not satisfy commonly accepted drug-like thresholds.

In addition to rule-based filtering, *Sanjeevini* 3.0 provides a Wiener index calculator as a topological descriptor to characterize molecular connectivity. In this module, molecules are represented as hydrogen-suppressed graphs, and a distance matrix is generated for all non-hydrogen atoms. The Wiener index is computed as the sum of the shortest-path distances, expressed in bond counts, between all atom pairs [40]. This descriptor reflects overall

molecular branching and connectivity and is used as an additional structural feature during early-stage filtering and screening.

2.9. StackTox for toxicity prediction of small molecules

StackTox is a machine learning-based toxicity prediction module integrated into the *Sanjeevini* 3.0 pipeline for early-stage safety assessment of small molecules. The module accepts molecular structures in SMILES format and predicts the probability of a compound being toxic or nontoxic, enabling the elimination of potentially unsafe candidates prior to downstream screening and docking. The model is trained on curated datasets derived from multiple publicly available sources, including PubChem repositories, FDA-approved drug datasets, KEGG, and other chemical databases, ensuring diverse representation of toxic and nontoxic chemical space. The StackTox model was specifically trained on the ToxBits dataset, comprising curated toxic and nontoxic molecules. To enable unbiased evaluation, an independent validation dataset (ToxBits-Val) was constructed using external sources such as Tox21 and human metabolite databases, ensuring no overlap with the training data.

Each molecule is represented using a compact set of physicochemical and topological descriptors calculated using RDKit. These descriptors capture key molecular properties including lipophilicity, polarity, molecular size, hydrogen-bonding characteristics, topological connectivity, and functional group composition, which are known to influence toxicity profiles. The current implementation predicts overall toxicity classification (toxic vs nontoxic), capturing signals associated with multiple toxicity endpoints such as hepatotoxicity, mutagenicity, and related adverse effects through its diverse training data. StackTox employs a stacked ensemble learning framework that integrates multiple base classifiers, including RF, Support Vector Machine (SVM), and Light Gradient Boosting Machine (LightGBM). The predictions from these base learners are combined using a meta-classifier (logistic regression) to generate the final toxicity prediction. This ensemble approach improves generalization and predictive robustness compared to individual models. The model performance was evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). The model demonstrated strong predictive performance, achieving an AUC-ROC of approximately 0.84 on the independent test set and ~0.72 on external validation datasets. The corresponding F1-score of ~0.76 further indicates balanced classification performance, supporting the robustness of the model across diverse chemical datasets. Benchmarking results further demonstrate that StackTox achieves competitive performance compared to existing toxicity prediction tools. The performance of the StackTox model across cross-validation and independent test datasets is summarized in Tables 2 and 3.

Within the *Sanjeevini* 3.0 workflow, StackTox serves as an early-stage filtering step, allowing users to prioritize compounds that satisfy both efficacy and safety criteria. The module is accessible through the web platform and supports high-throughput toxicity screening of large chemical libraries. The module can be accessed at https://scfbio-iitd.res.in/StackTox/.

2.10. Chemical database (BIMP) integrated in *Sanjeevini* 3.0

The in-house developed small molecule database available in the *Sanjeevini* 3.0 suite is BIMP. BIMP consolidates detailed information on 105,909 phytochemicals from 6,209 Indian medicinal plants. The plant list was compiled from major databases and literature covering Ayurveda, Siddha, Unani, and other traditional systems of medicine. Taxonomic data were retrieved from NCBI, and phytochemicals were curated from PubChem, IMMPAT [41], KNApSAcK [42], and additional sources, including their chemical identifiers (InChI, SMILES).

For structure preparation, 2D SMILES strings were standardized using RDKit. Protonation states were assigned assuming a physiological pH of 7.4 using RDKit's built-in pKa model. Chiral centers were retained as reported in the source databases; when stereochemistry was missing or ambiguous, RDKit was allowed to enumerate the most probable stereoisomers, after which duplicates and chemically unstable forms were removed.

Table 2. Cross-validation performance of individual base classifiers (Random Forest, LightGBM, and SVM) and the stacking ensemble model used in StackTox, reported as mean ± standard deviation across folds

Metric	RF	LGBM	SVM	Ensemble
AUC	0.832 ± 0.009	0.837 ± 0.008	0.810 ± 0.006	0.839 ± 0.008
Acc.	0.742 ± 0.019	0.745 ± 0.010	0.725 ± 0.010	0.751 ± 0.010
Prec.	0.713 ± 0.032	0.708 ± 0.015	0.682 ± 0.014	0.722 ± 0.016
Rec.	0.816 ± 0.030	0.834 ± 0.020	0.842 ± 0.011	0.820 ± 0.020
F1	0.760 ± 0.010	0.766 ± 0.008	0.754 ± 0.005	0.767 ± 0.008

Table 3. Performance of individual base classifiers and the stacking ensemble model on the independent test dataset (N = 2698), evaluated using standard classification metrics

Metric	Random Forest	LightGBM	SVM	Stacking ensemble
AUC-ROC	0.841	0.844	0.821	0.846
Accuracy	0.75	0.743	0.725	0.755
Precision	0.726	0.714	0.68	0.735
Recall	0.801	0.812	0.85	0.799
F1-Score	0.762	0.76	0.756	0.765

Three-dimensional conformers were generated using RDKit, followed by energy minimization with Balloon [43]. Physicochemical properties were computed with RDKit, and drug-likeness was assessed using standard filters (e.g., Lipinski and Veber rules).

BIMP also captures phytochemical-disease-gene associations, including over 30,000 phytochemical-disease links and hundreds of thousands of gene-target correlations. The web interface allows searching by plant name, family, BIMP ID, or physicochemical attributes, with integrated 2D/3D visualization. By combining curated phytochemical information with multiple screening and search capabilities, BIMP provides a comprehensive resource to support phytochemical-based drug discovery.

2.11. Monte Carlo-based docking protocol (ParDOCK+)

In instances where the binding cavity within the tertiary protein structure is already established, and comprehensive information regarding customized ligand molecules is at hand, the virtual screening step at the subsequent entry point may be bypassed. Instead, atomic-level docking and scoring of the tailor-made/selected ligands can be seamlessly executed through the pipeline, leveraging the ParDOCK+ module. It is noteworthy that this module operates in a fully automated, parallel processing mode for enhanced efficiency. ParDOCK+ specializes in conducting all-atom energy-based docking. It employs a Monte Carlo-based configurational search approach to ascertain the optimal ligand location and orientation within the active site. ParDOCK+ represents an improved version of the original ParDOCK module. This enhanced version has been equipped with specialized parameters to accommodate non-metallo and metalloprotein–ligand complexes. It incorporates specific parameters designed for ions such as Ca2+, Fe2+, Mg2+, Mn2+, and Zn2+ [44–46], thus broadening its utility for a wider range of molecular interactions. Within the domain of protein systems, several methodologies exist for modeling metal ions. Notable examples include the bonded model, nonbonded model, and the cationic dummy atom model [47]. The bonded model effectively addresses interactions between metals and ligands by specifying covalent bonds. Within this model, the potential energy function meticulously considers explicit bond and angle factors to precisely determine the associations between the metal and the atoms constituting the protein and ligand. However, it is important to note that this model is not without its limitations. It lacks the capability to replicate variations in coordination number and ligand exchange processes. Furthermore, it imposes restrictions on the overall conformational flexibility of the metal binding site. The cationic dummy model employs covalent bond theory to describe the interactions between the metal ion and the neighboring residues. This is achieved by assigning appropriate partial atomic charges and establishing a complex electrostatic force field. It is important to note that this method, owing to its reliance on a significant number of empirical parameters, demands a meticulous and sophisticated parameterization process. To circumvent the conformational limitations associated with the bonded model and the intricate parametrization process required by the cationic dummy model, an alternative nonbonded model has been incorporated into ParDOCK+. This nonbonded model includes electrostatic, van der Waals, and hydrophobic characteristics for the metal ion [48]. However, nonbonded models also come with their own set of limitations. Notably, they struggle to adequately account for the strong electrostatic interactions exhibited by divalent metal ions within the framework of the nonbonded model. ParDOCK+ ascertains the eight most favorable ligand poses within the binding cavity of the target protein. Following this, a minimization process consisting of 2500 steps is applied to the complex. Additionally, work is currently in progress to extend the docking module with machine learning and deep learning-based flexible docking capabilities.

2.12. Binding affinity estimation (BAPPL+)

The scoring module serves as the next stage in the workflow, where docked protein–ligand complexes are utilized to estimate binding affinity within the active site. The prioritization of minimized complexes is performed using the BAPPL+ scoring function, which provides the top-ranked ligand poses generated by the ParDOCK+ module.

BAPPL+ is a computationally efficient scoring method designed for both non-metallo and metalloprotein–ligand systems. The method estimates binding free energy by integrating key physicochemical interaction components, including electrostatic (ΔEele), van der Waals (ΔEvdw), hydrophobic (ΔEhyp), and conformational entropy (−TΔSconf) contributions.

The predicted binding free energy is given by:

$$\Delta G^{\circ}\text{pred} = \Delta\text{Eele} + \Delta\text{Evdw} + \Delta\text{Ehyp} - T\Delta\text{Sconf} \quad (1)$$

where ΔEele represents electrostatic interactions, ΔEvdw accounts for van der Waals forces, ΔEhyp corresponds to hydrophobic interactions, and TΔSconf denotes the loss of conformational entropy upon ligand binding. A key advantage of BAPPL+ is its ability to accurately model metalloprotein–ligand interactions through the incorporation of parameters for commonly occurring metal ions, including Zn^{2+}, Mg^{2+}, Ca^{2+}, Mn^{2+}, and Fe^{3+}. This enables improved prediction accuracy for metal-containing binding sites, which are often challenging for conventional scoring functions. In addition, BAPPL+ integrates a machine learning-based regression model to enhance predictive performance. In this work, an RF algorithm is employed to capture nonlinear relationships between molecular descriptors and binding affinity, improving generalization across diverse protein–ligand systems. The performance of BAPPL+ has been evaluated using benchmark datasets, demonstrating strong agreement with experimental binding affinities, with a Pearson correlation coefficient of approximately 0.76. The module can be accessed at https://scfbio-iitd.res.in/bappl+/.

2.13. Molecular dynamics simulations

Molecular dynamics (MD) simulations are recommended as a downstream analysis step following docking to move beyond static protein–ligand structures and assess dynamic stability and ensemble-averaged properties. MD allows users to examine whether a docked ligand remains stably bound over time, to capture conformational flexibility of the protein–ligand complex, and to account for explicit solvent, salt, and temperature effects. Such dynamic evaluation provides a more realistic description of binding behavior compared to single-structure docking scores. MD trajectories generated by users can be further analyzed using post facto free-energy methods, including MMGBSA, MMPBSA [49, 50], and MMBAPPL+, enabling estimation of binding free energies as ensemble averages. These analyses can be used to assess improvements in ranking and correlation with experimentally reported binding affinities, thereby providing thermodynamically informed validation of docking results.

Table 4. Contextual comparison of *Sanjeevini* 3.0 machine learning and docking-based modules with representative state-of-the-art tools. Reported metrics are taken from respective literature sources and may not be directly comparable due to differences in datasets, evaluation protocols, and computational environments

Category	Tool name	Method	Key metric (accuracy)[1]	Hardware	Screening time (per one million compounds)
ML/DL-based(regression)	SEARCH-ML	ML Ensemble (CPU-Optimized)	0.72	CPU	~0.15 min (10 s)
	RASPD+	ERF, RF, DNN, KNN, SVR, eSVR, LR	0.52	CPU	~30 min
	GraphDTA [51]	Graph Neural Network	0.67	GPU	~45 min
	KDEEP [52]	3D Convolutional Neural Network	0.67	GPU	~90 min
	OnionNet-2 [38]	3D Convolutional Neural Network	0.74	GPU	N/A
ML/DL-based (classifier)	DeepDock [53]	Deep Neural Network (DNN)	AUPR = 0.38[2]	GPU	N/A
	Deep Docking (D2) [54]	DL-QSAR (Docking Accelerator)	EF = ~6000x[2]	GPU	N/A

The simulations can be carried out using the computing resources available at the SCFBio, IIT Delhi. Commonly used MD packages such as AMBER, GROMACS, and NAMD are already installed on the system, and users can run simulations for up to 360 h.

A benchmarking summary of the machine learning and docking modules, including the methods used, accuracy metrics, hardware configuration, and screening time in comparison with existing tools, is provided in Table 4.

Overall, the *Sanjeevini* 3.0 website provides users with sample input files and comprehensive online tutorials for each entry point. Users have the option to use default or customized settings. Each query is assigned a unique Job ID, which can be used to access results or check the job status.

3. Results and Discussion

3.1. Validation of the docking and scoring modules

To appraise the docking and scoring processes, the PDBbind database furnishes specialized datasets encompassing complexes distinguished by experimentally verified binding affinities. For the validation of ParDOCK+, a dataset comprising 483 non-metalloprotein complexes was selected. These complexes were chosen based on their availability in the refined data of PDBbind, with known binding affinities for reference. The selection of these complexes was conducted randomly regarding protein families, yet with a key criterion being well-resolved structures, all of which exhibit a resolution of less than 3.0 Å. The selected proteins and ligands were downloaded from the RCSB Protein Data Bank. The evaluation of the performance of ParDOCK+ was based on the comparison of experimental and predicted binding affinities as well as the root mean square deviation (RMSD) of the structure between the experimental and top-docked complex. The validation dataset for non-metalloproteins, featuring PDB IDs along with experimental and projected binding free energies, as well as RMSD values, is lucidly detailed in Supplementary Table S1. It is worth noting that through this protocol, an impressive correlation of r = 0.812 has been established between the predicted and experimental binding affinities, as illustrated in Figure 4(A). The average RMSD value, indicating the disparity between experimental and docked complexes, stands at 0.39 Å. Approximately 176 complexes exhibit predicted binding energies that deviate by more than 2 kcal/mol. The range of rest is from 2 to 0. There are 168 complexes in total with experimental value deviations below 1 kcal/mol, as depicted in Figure 5(A).

For validating metalloprotein complexes involving any of the five ions (Ca2+, Fe2+, Mg2+, Mn2+, and Zn2+), an extensive dataset comprising a total of 1157 such complexes was obtained from the RCSB database. This dataset includes diverse categories, namely, 352 calcium-containing metalloproteins, 19 iron-containing metalloproteins, 245 magnesium-containing metalloproteins, 41 manganese-containing metalloproteins, and 500 zinc-containing metalloproteins. The correlation coefficients between predicted and experimental binding affinities for metalloprotein–ligand compounds are as follows: 0.740 for Ca2+, 0.520 for Fe2+, 0.720 for Mg2+, 0.716 for Mn2+, and 0.753 for Zn2+, as visually represented in Figure 4. Table 5 provides a detailed presentation of key metrics, including the RMSD measuring structural alignment between the experimental and top-docked complex, the disparity between predicted and experimental affinities, and the calculated average deviation value. Figure 5 illustrates a prominent pattern of deviation, predominantly observed in cases of either very high affinity (exceeding 11 kcal/mol) or very low affinity (below 4 kcal/mol). Notably, the affinity prediction demonstrates a much higher degree of accuracy within the 4–10 kcal/mol range. The validation dataset for metalloproteins, featuring PDB IDs along with experimental and projected binding free energies, as well as RMSD values, is given in Supplementary Table S2.

The duration of execution per job was meticulously monitored, utilizing an 8-processor Intel(R) Xeon(R) cluster, encompassing a dataset of 1712 complexes. The comprehensive analysis reveals an average runtime of 6.29 min. It is noteworthy that the CPU runtime exhibits variability, ranging between 2 and 13 min, with primary dependence on the dimensions of the ligand

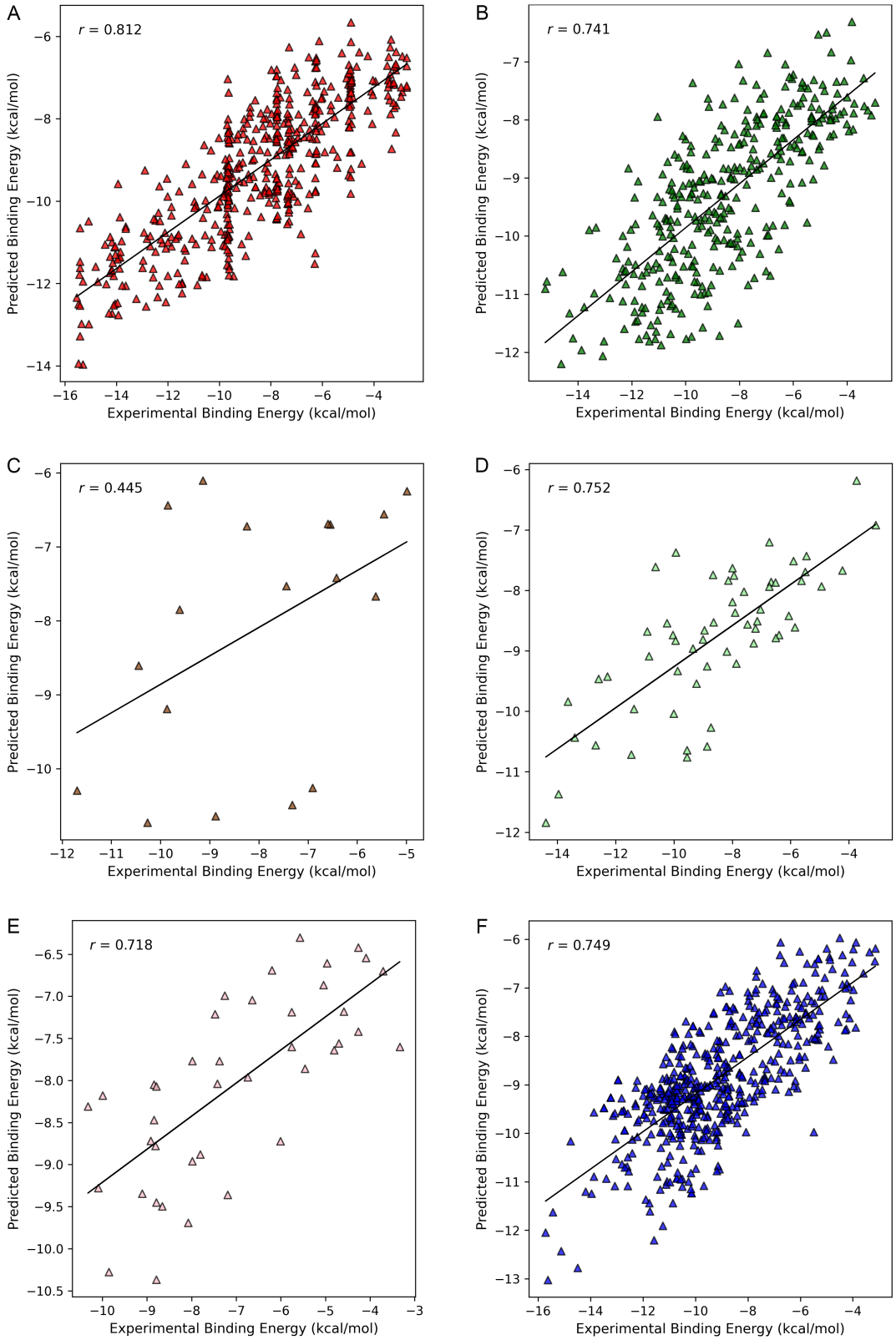

Figure 4. Correlation between predicted and experimental binding affinities for (A) non-metalloprotein–ligand complexes, (B) Ca2+ containing metalloprotein–ligand complexes, (C) Fe2+ containing metalloprotein–ligand complexes, (D) Mg2+ containing metalloprotein–ligand complexes, (E) Mn2+ containing metalloprotein–ligand complexes, and (F) Zn2+ containing metalloprotein–ligand complexes

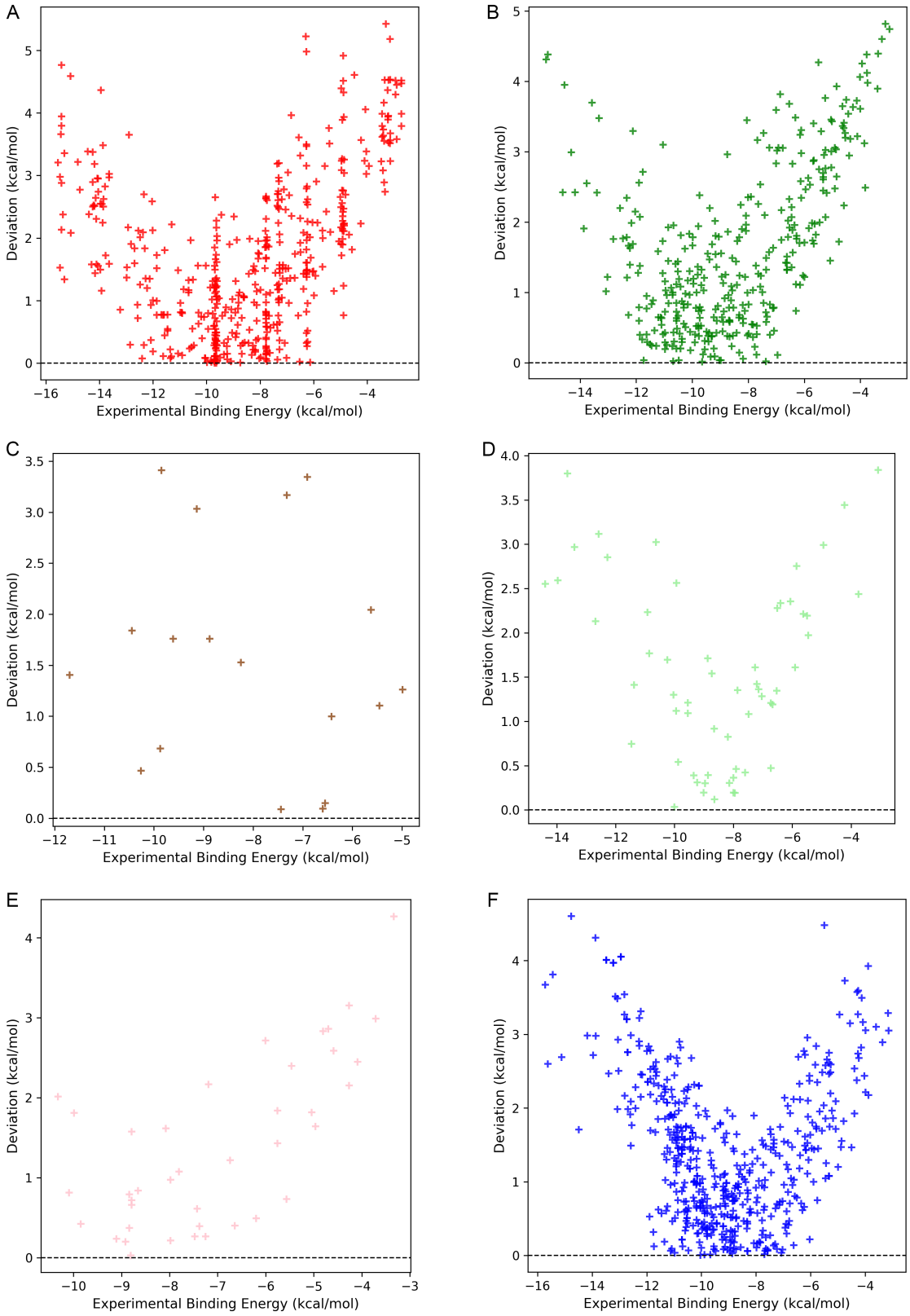

Figure 5. Deviation of predicted binding energies from experimental values plotted against experimental binding energy (kcal/mol) for (A) non-metalloprotein–ligand complexes, (B) Ca2+ containing metalloprotein–ligand complexes, (C) Fe2+ containing metalloprotein–ligand complexes, (D) Mg2+ containing metalloprotein–ligand complexes, (E) Mn2+ containing metalloprotein–ligand complexes, and (F) Zn2+ containing metalloprotein–ligand complexes

Table 5. Validation summary of metalloprotein–ligand complexes for different metal ions, reporting structural accuracy (RMSD), deviation of predicted binding affinities from experimental values, and average error (kcal/mol)

Metal ions in protein	Average RMSD	Number of complexes deviating by more than 2 kcal/mol from experimental binding energies	Number of complexes deviating by less than 1 kcal/mol from experimental binding energies	Average deviation (in kcal/mol)
Ca^{2+}	0.407	107	135	1.59
Fe^{2+}	0.390	6	6	1.62
Mg^{2+}	0.409	81	82	1.57
Mn^{2+}	0.382	12	20	1.38
Zn^{2+}	0.563	124	213	1.38

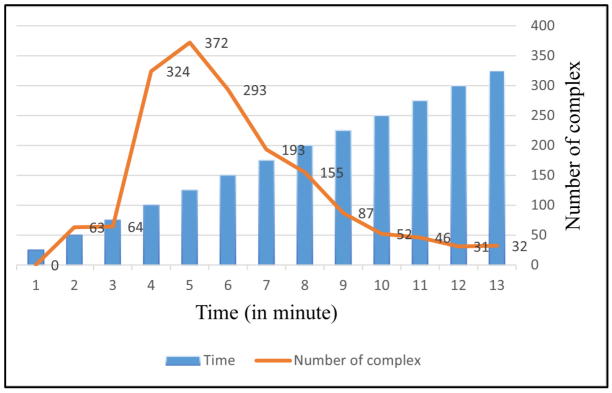

Figure 6. CPU runtime distribution for complexes

molecule. The runtime for most of the complexes falls within the 4–6-min range, as visually represented in Figure 6. Conversely, the size of the protein molecule exerts a comparatively minimal influence on the runtime of ParDOCK+. It is imperative to acknowledge that the actual runtime may experience variations based on the availability of free server nodes.

3.2. Validation of complete pipeline with FDA-approved drugs

To evaluate its applicability across diverse therapeutic areas, the *Sanjeevini* 3.0 pipeline was validated using protein targets associated with cancer, cardiovascular, pulmonary, autoimmune, neurological, genetic, bacterial, and viral diseases. The selected targets are not consistently distributed among disease groups, as inclusion was mostly based on the availability of high-resolution protein structures with co-crystallized ligands in the PDB. For this, 120 FDA-approved drugs and 126 pharmacologically active targets were carefully coupled to create 194 protein–ligand complexes for research and evaluation. The RCSB database provided experimentally generated PDB structures of target proteins. Additionally, FDA-approved drugs associated with these target proteins were acquired from the DrugBank [28] and PubChem databases [55, 56] in 3D format. Tertiary protein structures were fed to the active site prediction program, which identified probable binding holes in 30 s using physicochemical parameters. The top 10 binding cavities of each target were then virtually tested against an FDA-approved pre-parameterized RASPD+ module library. One active site of the target protein can be screened against an FDA-approved library in 1 min to determine ligand–target protein binding affinity. RASPD+ predicts target protein hits well despite its computational speed. The FDA-approved medications associated with the input target proteins were usually filtered out as hits with good binding affinities (Figure 6). The protein and drug tertiary structures were submitted to ParDOCK+ with default parameters to determine drug molecule positions in the binding pocket. Depending on receptor and drug candidate size, this upgraded ParDOCK+ module completed docking operations in 2–13 min. In 1 min, the BAPPL+ module prioritized drug poses. To classify FDA-approved medications as hit molecules, we examined their rank and energy values. Rediscovery was the identification of the FDA-approved medicine that matches a target among the top-ranked projected candidates based on binding energy (BAPPL+ score) and ranking criteria after docking. The rediscovery percentage was the ratio of known drug–target pairs found to evaluated complexes. After docking, complexes with binding energies above −5 kcal/mol were deemed non-hits (Table S3). Because docking scores over this range imply weak binding, this cutoff was used. It easily separates likely binders from non-binders. Figure 7 shows that *Sanjeevini* 3.0 modules found previously known hit compounds for the target protein

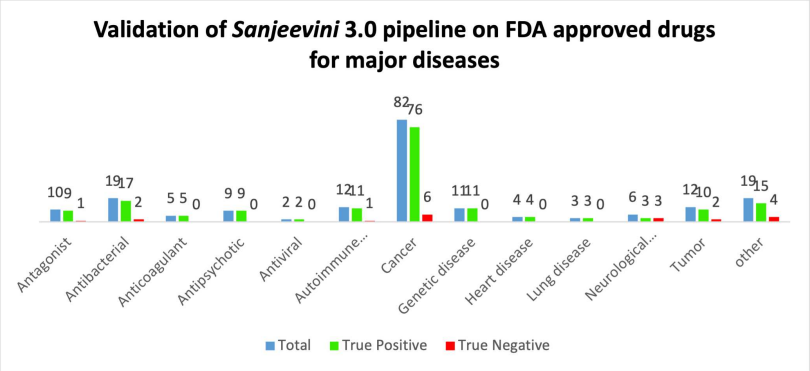

Figure 7. A histogram plot showing hits (true positives) and non-hits (true negatives) for various diseases using a cutoff of −5 kcal/mol

in 90% of cases. This strong result proves the software suite's accuracy and reliability. This retrospective evaluation validates, but it is based on known drug–target connections and may not accurately predict future performance.

Structure-based drug discovery procedures require successful rediscovery of known ligands as a validation requirement, which matches these results. The strong integrated screening, docking, and scoring system allows *Sanjeevini* 3.0 to identify 90% of FDA-approved medications. The new *Sanjeevini* framework includes machine learning-based screening, enhanced metalloprotein–ligand interaction treatment, and early-stage toxicity evaluation, overcoming various drawbacks of previous implementations. These improvements coincide with recent CADD advancements that use hybrid physics-based and data-driven approaches to improve forecast accuracy and scalability.

Sanjeevini 3.0 integrates physics-based potentials with advanced machine learning techniques to provide practically all protein-targeted drug design modules. In addition to ML-augmented screening, docking, and scoring, this edition includes a web-accessible toxicity prediction module for early-stage small molecule safety assessment before downstream screening. The usability-optimized platform lets researchers undertake sophisticated computational procedures with minimal technological overhead while assuring methodological consistency through curated chemical libraries. When tested against non-metalloprotein–ligand complexes, the pipeline correlated above 0.812 with experimental binding energies. This paradigm is based on translational relevance; prior versions were used to uncover antiviral inhibitors for viral proteases, hepatitis B, and chikungunya, cancer targets, antimalarial lead optimization, and neurological enzymes [6].

The existing framework has limits despite its outstanding performance. The literature-reported performance metrics used to compare methods may not be directly benchmarking under identical datasets and computational conditions. Therefore, these comparisons should be interpreted in a broader contextual framework rather than as strict head-to-head evaluations. The system uses static protein structure docking and scoring, which may not capture conformational flexibility and dynamic binding effects. MD simulations are indicated as a downstream step but not yet in the automated pipeline. The diversity and quality of training datasets affect the performance of machine learning-based screening modules, which may affect generalization to novel chemical spaces. Current computational models use approximations to treat metalloprotein–ligand interactions, despite improvements. The StackTox module only classifies binary toxicity and does not resolve individual endpoints. These constraints suggest methodological improvements.

4. Conclusion

Sanjeevini 3.0's validation results show that it can dependably identify promising lead compounds and rediscover FDA-approved target protein medicines, enabling its use in structure-based drug discovery workflows. The platform prioritizes candidate compounds efficiently and methodically by merging machine learning-based screening, physics-based docking and scoring, extended chemical libraries, and early-stage toxicity prediction. This work improves metalloprotein–ligand interactions, data-driven screening, and pharmacokinetic and safety filtering in *Sanjeevini*, addressing several limitations of previous implementations. *Sanjeevini* 3.0 is a comprehensive and computationally efficient platform for conventional and data-driven drug development, helping molecular insights become candidate treatments. For detailed user manuals on modules and workflows, see https://scfbio-iitd.res.in/Sanjeevini/User-Mannuals.php.

Acknowledgment

Funding to SCFBio, IIT Delhi, from the National Supercomputing Mission administered by DEITY, DST & CDAC of the Govt. of India, and CoE funding from the Department of Biotechnology, Govt. of India, are gratefully acknowledged. Several rounds of validations of *Sanjeevini* 3.0 modules conducted by Ms. Akshata Hegde, Ms. Smriti Pranjal, Ms. Bhumika Singh, Mr. Devendra Prajapat, Dr. Amita Pathak, Dr. Ruchika Bhat, and several project trainees are also sincerely acknowledged.

Funding Support

The authors gratefully acknowledge support to SCFBio, IIT Delhi, from NSM, MEITY & DST, Govt. of India, and DBT, Govt. of India.

Ethical Statement

This study did not involve experiments on human participants or animals performed by the authors. All datasets used in this work were obtained from publicly available databases and resources collected under the respective ethical approvals and guidelines of the original studies.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The datasets used and/or generated during the current study are available from the corresponding authors upon reasonable request. The PDBbind datasets used for training and validation are publicly available at https://www.pdbbind-plus.org.cn/. The chemical libraries used for screening, including ZINC (https://zinc15.docking.org/), DrugBank (https://go.drugbank.com/), and FDA-approved compounds (https://go.drugbank.com/), are available from their respective public repositories.

The StackTox training (ToxBits) and external validation (ToxBits-Val) datasets can be provided by the corresponding authors upon request.

Author Contribution Statement

Dheeraj Kumar Chaurasia: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Aman Sharma:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Madhvi Mishra:** Software, Validation, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Akanksha Kesharwani:** Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft. **Raushan Anjum:** Software, Validation, Investigation, Data curation. **Shashank Shekhar:** Software, Resources,

Project administration. **Aditya Mittal:** Software, Resources, Project administration, Funding acquisition. **B. Jayaram:** Conceptualization, Methodology, Software, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Supplementary Information

The supplementary table for this article can be found at https://doi.org/10.47852/bonviewMEDIN62028879.

References

[1] Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á., & Garcia-Fandino, R. (2023). The role of AI in drug discovery: Challenges, opportunities, and strategies. *Pharmaceuticals*, *16*(6), 891. https://doi.org/10.3390/ph16060891

[2] Yadav, S., Singh, A., Singhal, R., & Yadav, J. P. (2024). Revolutionizing drug discovery: The impact of artificial intelligence on advancements in pharmacology and the pharmaceutical industry. *Intelligent Pharmacy*, *2*(3), 367–380. https://doi.org/10.1016/j.ipha.2024.02.009

[3] Kumar, A., Gupta, S. K., & Kim, S. (2025). AI-driven drug discovery using a context-aware hybrid model to optimize drug-target interactions. *Scientific Reports*, *15*(1), 35719. https://doi.org/10.1038/s41598-025-19593-4

[4] Chakraborty, C., Bhattacharya, M., Lee, S.-S., Wen, Z.-H., & Lo, Y.-H. (2024). The changing scenario of drug discovery using AI to deep learning: Recent advancement, success stories, collaborations, and challenges. *Molecular Therapy Nucleic Acids*, *35*(3), 102295. https://doi.org/10.1016/j.omtn.2024.102295

[5] Roney, M., & Mohd Aluwi, M. F. F. (2024). The importance of in-silico studies in drug discovery. *Intelligent Pharmacy*, *2*(4), 578–579. https://doi.org/10.1016/j.ipha.2024.01.010

[6] Singh, A., Shekhar, S., & Jayaram, B. (2021). Cadd: Some success stories from Sanjeevini and the way forward. In S. K. Singh (ed) (Ed.), *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design* (pp. 1–18). Springer Singapore. https://doi.org/10.1007/978-981-15-8936-2_1

[7] Niazi, S. K., & Mariam, Z. (2023). Computer-aided drug design and drug discovery: A prospective analysis. *Pharmaceuticals*, *17*(1), 22. https://doi.org/10.3390/ph17010022

[8] Cui, M. (2024). The application of computer-aided drug design technology in drug development. In *Proceedings of the 2024 5th International Symposium on Artificial Intelligence for Medicine Science*, 880–886. https://doi.org/10.1145/3706890.3707041

[9] Sadybekov, A. V., & Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature*, *616*(7958), 673–685. https://doi.org/10.1038/s41586-023-05905-z

[10] King, E., Aitchison, E., Li, H., & Luo, R. (2021). Recent developments in free energy calculations for drug discovery. *Frontiers in Molecular Biosciences*, *8*, 712085. https://doi.org/10.3389/fmolb.2021.712085

[11] Bhat, R., Kaushik, R., Singh, A., DasGupta, D., Jayaraj, A., Soni, A., . . . , & Jayaram, B. (2020). A comprehensive automated computer-aided discovery pipeline from genomes to hit molecules. *Chemical Engineering Science*, *222*, 115711. https://doi.org/10.1016/j.ces.2020.115711

[12] Garg, P., Malhotra, J., Kulkarni, P., Horne, D., Salgia, R., & Singhal, S. S. (2024). Emerging therapeutic strategies to overcome drug resistance in cancer cells. *Cancers*, *16*(13), 2478. https://doi.org/10.3390/cancers16132478

[13] Li, B., Wang, Z., Liu, Z., Tao, Y., Sha, C., He, M., & Li, X. (2024). DrugMetric: Quantitative drug-likeness scoring based on chemical space distance. *Briefings in Bioinformatics*, *25*(4), bbae321. https://doi.org/10.1093/bib/bbae321

[14] Burley, S. K., Bhatt, R., Bhikadiya, C., Bi, C., Biester, A., Biswas, P., . . . , & Zardecki, C. (2025). Updated resources for exploring experimentally-determined PDB structures and computed structure models at the RCSB protein data bank. *Nucleic Acids Research*, *53*(D1), D564–D574. https://doi.org/10.1093/nar/gkae1091

[15] Jia, Z. C., Yang, X., Wu, Y. K., Li, M., Das, D., Chen, M. X., & Wu, J. (2024). The art of finding the right drug target: Emerging methods and strategies. *Pharmacological Reviews*, *76*(5), 896–914. https://doi.org/10.1124/pharmrev.123.001028

[16] Palermo, G., Spinello, A., Saha, A., & Magistrato, A. (2021). Frontiers of metal-coordinating drug design. *Expert Opinion on Drug Discovery*, *16*(5), 497–511. https://doi.org/10.1080/17460441.2021.1851188

[17] Majorek, K. A., Gucwa, M., Murzyn, K., & Minor, W. (2024). Metal ions in biomedically relevant macromolecular structures. *Frontiers in Chemistry*, *12*, 1426211. https://doi.org/10.3389/fchem.2024.1426211

[18] Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2022). Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv*. https://doi.org/10.48550/ARXIV.2210.01776

[19] Stärk, H., Ganea, O., Pattanaik, L., Barzilay, R., & Jaakkola, T. (2022). Equibind: Geometric deep learning for drug binding structure prediction. *arXiv Preprint: 2202.05146*.

[20] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

[21] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. https://doi.org/10.1016/0022-2836(82)90153-X

[22] Smith, A. E., & Lindner, H. J. (1991). -scf-molecular mechanics PIMM: Formulation, parameters, applications. *Journal of Computer-Aided Molecular Design*, *5*(3), 235–262. https://doi.org/10.1007/BF00124341

[23] Meier, R., Pippel, M., Brandt, F., Sippl, W., & Baldauf, C. (2010). paradocks: A framework for molecular docking with population-based metaheuristics. *Journal of Chemical Information and Modeling*, *50*(5), 879–889. https://doi.org/10.1021/ci900467x

[24] Yu, Y., Wang, R., & Teo, R. D. (2022). Machine learning approaches for metalloproteins. *Molecules*, *27*(4), 1277. https://doi.org/10.3390/molecules27041277

[25] Huang, K., Xiao, C., Glass, L. M., & Sun, J. (2021). MolTrans: Molecular Interaction Transformer for drug–target interaction prediction. *Bioinformatics*, *37*(6), 830–836. https://doi.org/10.1093/bioinformatics/btaa880

[26] Tingle, B. I., Tang, K. G., Castanon, M., Gutierrez, J. J., Khurelbaatar, M., Dandarchuluun, C., . . . , & Irwin, J. J. (2023). Zinc-22—a free multi-billion-scale database of tangible compounds for ligand discovery. *Journal of Chemical Information and Modeling*, *63*(4), 1166–1176. https://doi.org/10.1021/acs.jcim.2c01253

[27] Chaurasia, D. K., Anjum, R., Sharma, A., Mishra, M., Shekhar, S., Patel, A. K., . . . , & Jayaram, B. (2026). BIMP: Unveiling the phytochemical richness of Indian medicinal plants as potential therapeutic agents. In A. K. Saxena & A. Saxena (Eds.), *Global Trends in Health, Technology and Management II* (pp. 283–298). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-12320-6_16

[28] Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., . . . , & Wishart, D. S. (2024). DrugBank 6.0: The DrugBank Knowledgebase for 2024. *Nucleic Acids Research*, *52*(D1), D1265–D1275. https://doi.org/10.1093/nar/gkad976

[29] Kesharwani, A., Chaurasia, D. K., & Katara, P. (2022). Repurposing of FDA approved drugs and their validation against potential drug targets for Salmonella enterica through molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics*, *40*(14), 6255–6271. https://doi.org/10.1080/07391102.2021.1880482

[30] Orlandi, M., Geng, Y., Macchiagodena, M., Pagliai, M., & Procacci, P. (2025). Solvation free energies of drug-like molecules via fast growth in an explicit solvent: Assessment of the am1-bcc, resp/hf/6–31g*, resp-qm/mm, and abcg2 fixed-charge approaches. *Journal of Chemical Theory and Computation*, *21*(16), 7977–7990. https://doi.org/10.1021/acs.jctc.5c00749

[31] Wang, Y., Pulido, I., Takaba, K., Kaminow, B., Scheen, J., Wang, L., & Chodera, J. D. (2024). Espalomacharge: Machine learning-enabled ultrafast partial charge assignment. *The Journal of Physical Chemistry A*, *128*(20), 4160–4167. https://doi.org/10.1021/acs.jpca.4c01287

[32] Guo, D., Feng, L., Shi, C., Cao, L., Li, Y., Wang, Y., & Xu, X. (2022). Vappd: Visual analysis of protein pocket dynamics. *Applied Sciences*, *12*(20), 10465. https://doi.org/10.3390/app122010465

[33] Darlami, J., & Sharma, S. (2024). The role of physicochemical and topological parameters in drug design. *Frontiers in Drug Discovery*, *4*, 1424402. https://doi.org/10.3389/fddsv.2024.1424402

[34] Li, J., Guan, X., Zhang, O., Sun, K., Wang, Y., Bagni, D., & Head-Gordon, T. (2026). Leak proof PDBbind: A reorganized data set of protein–ligand complexes for more generalizable binding affinity prediction. *The Journal of Physical Chemistry B*, *130*(2), 730–740. https://doi.org/10.1021/acs.jpcb.5c08598

[35] Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. *Technologies*, *13*(3), 88. https://doi.org/10.3390/technologies13030088

[36] Florek, P., & Zagdański, A. (2023). Benchmarking state-of-the-art gradient boosting algorithms for classification. *arXiv*. https://doi.org/10.48550/ARXIV.2305.17094

[37] Wen, Y., Guo, R., Duan, Z., Tong, Y., Tang, X., Pan, T., & Fu, C. (2026). Machine learning model optimization with optuna for accurate prediction of strength and crack behavior in prestressed concrete beams. *Scientific Reports*, *16*(1), 5822. https://doi.org/10.1038/s41598-026-36692-y

[38] Wang, Z., Zheng, L., Liu, Y., Qu, Y., Li, Y. Q., Zhao, M., . . . , & Li, W. (2021). Onionnet-2: A convolutional neural network model for predicting protein-ligand binding affinity based on residue-atom contacting shells. *Frontiers in Chemistry*, *9*, 753002. https://doi.org/10.3389/fchem.2021.753002

[39] Mandujano-Lázaro, G., Torres-Rojas, M. F., Ramírez-Moreno, E., & Marchat, L. A. (2025). Virtual screening combined with molecular docking for the identification of new anti-adipogenic compounds. *Science Progress*, *108*(1), 00368504251320313. https://doi.org/10.1177/00368504251320313

[40] Ashraf, S., Imran, M., Bokhary, S. A. U. H., & Akhter, S. (2022). The Wiener index, degree distance index and Gutman index of composite hypergraphs and sunflower hypergraphs. *Heliyon*, *8*(12), e12382. https://doi.org/10.1016/j.heliyon.2022.e12382

[41] Vivek-Ananth, R. P., Mohanraj, K., Sahoo, A. K., & Samal, A. (2023). Imppat 2. 0: An enhanced and expanded phytochemical atlas of indian medicinal plants. *ACS Omega*, *8*(9), 8827–8845. https://doi.org/10.1021/acsomega.3c00156

[42] Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., . . . , & Kanaya, S. (2012). Knapsack family databases: Integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology*, *53*(2), e1–e1. https://doi.org/10.1093/pcp/pcr165

[43] Schmid, S. P., Seng, H., Kläy, T., & Jorner, K. (2026). Rapid generation of transition-state conformer ensembles via constrained distance geometry. *Journal of Chemical Information and Modeling*, *66*(5), 2777–2790. https://doi.org/10.1021/acs.jcim.5c02794

[44] Ye, N., Zhou, F., Liang, X., Chai, H., Fan, J., Li, B., & Zhang, J. (2022). A comprehensive review of computation-based metal-binding prediction approaches at the residue level. *BioMed Research International*, *2022*(1), 8965712. https://doi.org/10.1155/2022/8965712

[45] Clemente, C. M., Prieto, J. M., & Martí, M. (2024). Unlocking precision docking for metalloproteins. *Journal of Chemical Information and Modeling*, *64*(5), 1581–1592. https://doi.org/10.1021/acs.jcim.3c01853

[46] Jiang, D., Ye, Z., Hsieh, C. Y., Yang, Z., Zhang, X., Kang, Y., . . . , & Hou, T. (2023). MetalProGNet: A structure-based deep graph model for metalloprotein–ligand interaction predictions. *Chemical Science*, *14*(8), 2054–2069. https://doi.org/10.1039/D2SC06576B

[47] Volkenandt, S., & Imhof, P. (2023). Comparison of empirical Zn2+ models in protein–DNA complexes. *Biophysica*, *3*(1), 214–230. https://doi.org/10.3390/biophysica3010014

[48] Piskorz, T. K., Lee, B., Zhan, S., & Duarte, F. (2024). metallicious: Automated force-field parameterization of covalently bound metals for supramolecular structures. *Journal of Chemical Theory and Computation*, *20*(20), 9060–9071. https://doi.org/10.1021/acs.jctc.4c00850

[49] Jiang, D., Du, H., Zhao, H., Deng, Y., Wu, Z., Wang, J., . . . , & Hsieh, C. Y. (2024). Assessing the performance of MM/PBSA and MM/GBSA methods. 10. Prediction reliability of binding affinities and binding poses for RNA–ligand complexes. *Physical Chemistry Chemical Physics*, *26*(13), 10323–10335. https://doi.org/10.1039/D3CP04366E

[50] Roux, B., & Chipot, C. (2024). Editorial guidelines for computational studies of ligand binding using MM/PBSA and

MM/GBSA approximations wisely. *The Journal of Physical Chemistry B*, *128*(49), 12027–12029. https://doi.org/10.1021/acs.jpcb.4c06614

[51] Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, *37*(8), 1140–1147. https://doi.org/10.1093/bioinformatics/btaa921

[52] Jiménez, J., Škalič, M., Martínez-Rosell, G., & De Fabritiis, G. (2018). kdeep: Protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of Chemical Information and Modeling*, *58*(2), 287–296. https://doi.org/10.1021/acs.jcim.7b00650

[53] Liao, Z., You, R., Huang, X., Yao, X., Huang, T., & Zhu, S. (2019). Deepdock: Enhancing ligand-protein interaction prediction by a combination of ligand and structure information. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 311–317. https://doi.org/10.1109/BIBM47256.2019.8983365

[54] Gentile, F., Agrawal, V., Hsing, M., Ton, A.-T., Ban, F., Norinder, U., . . . , & Cherkasov, A. (2020). Deep docking: A deep learning platform for augmentation of structure based drug discovery. *ACS Central Science*, *6*(6), 939–949. https://doi.org/10.1021/acscentsci.0c00229

[55] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., . . . , & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*, *51*(D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956

[56] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., . . . , & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

How to Cite: Chaurasia, D. K., Sharma, A., Mishra, M., Kesharwani, A., Anjum, R., Shekhar, S., …, & Jayaram, B. (2026). *Sanjeevini* 3.0: An Enhanced Comprehensive Automated Web Server for Computer-Aided Drug Design. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN62028879