ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Machine Learning, Molecular Docking, and Dynamics-Based Computational Identification of Potential Inhibitors against Lung Cancer

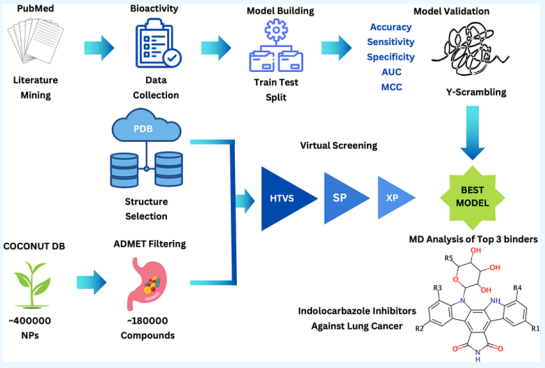
Agneesh Pratim Das, Puniti Mathur, and Subhash M. Agarwal*

Cite This: *ACS Omega* 2024, 9, 4528−4539

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Lung cancer is the most prevalent cause of cancer deaths worldwide. However, its treatment faces a significant hurdle due to the development of resistance. Phytomolecules are an important source of new chemical entities due to their rich chemical diversity. Therefore, a machine learning (ML) model was developed to computationally identify potential inhibitors using a curated data set of 649 phytomolecules with inhibitory activity against lung cancer cell lines. Four distinct ML approaches, including *k*-nearest neighbor, random forest, support vector machine, and extreme gradient boosting, were used in conjugation with MACCS and Morgan2 fingerprints to generate the models. It was observed that the random forest model developed by using the MACCS fingerprint shows the best performance. To further explore the chemical space and feature importance, *k*-means clustering, t-SNE analysis, and mean decrease in impurity had been calculated. Simultaneously, ∼400 000 natural products (NPs) retrieved from the COCONUT database were filtered for pharmacokinetic properties and taken for a multistep screening using docking against epidermal growth factor receptor (EGFR) mutant, a therapeutic drug target of lung cancer. Thereafter, the best-performing random forest model was used to predict the antilung cancer potential of the NPs having binding affinity better than the cocrystal ligand. This allowed the identification of 205 potential inhibitors, wherein the molecules with an indolocarbazole scaffold were enriched in top-scoring molecules. The top three indolocarbazole molecules with the lowest binding energy were further evaluated through 100 ns molecular dynamics (MD) simulations, which suggested that these molecules are strong binders. Also, structural similarity analysis against known drugs revealed that these NPs are similar to staurosporine, which demonstrates potent and selective activity against EGFR mutants. Thereby, the consensus analysis employing ML, molecular docking, and dynamics revealed that the molecules having an indolocarbazole scaffold are the most promising NPs that can act as potential inhibitors against lung cancer.

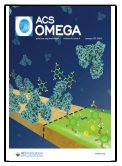


INTRODUCTION

Cancer is one of the major causes of death globally and the number of patients dying due to it is continuously rising every year.[1] Among the different types of cancers, lung cancer is the most prevalent and the primary cause of cancer-related deaths.[2] However, despite several chemotherapeutic treatment options, lung cancer remains a major concern due to drug resistance. Thus, there is a need for the identification of new and potent antilung cancer compounds.[3] In several nations globally, medicinal plant-based remedies are popularly used.[4,5] As per the latest and most comprehensive natural product drug discovery review published by Newman and Cragg in 2020, around 33.5% of the small-molecule anticancer drugs developed during 1981−2019 were either direct natural products (NPs) or their derivatives.[6] This establishes that NPs continue to be a primary source for the development of new therapeutic anticancer drugs.

To find novel bioactive molecules against cancer drug targets, the conventional structural approaches are extensively used for in silico screening of chemical databases.[5,7,8] Nonetheless, given the advent of advanced computational techniques like machine learning (ML), it has found its application in various scientific problems.[9−11] The ML classification models work on the preamble that the structures of biologically active molecules have common features that are responsible for their bioactivity.[11,12] So, to accelerate the drug discovery process, an integrated approach utilizing both ML prediction models and structure-based virtual screening along with molecular dynamics must be implemented to computationally identify molecules that can potentially inhibit lung cancer drug targets.

Received: September 23, 2023
Revised: November 29, 2023
Accepted: November 30, 2023
Published: January 17, 2024



Therefore, in the present study, the reported biological activity of phytomolecules against lung cancer cell lines was extracted from the literature by text mining for building ML-based prediction models. Multiple ML models were then developed by using a combination of different ML algorithms and molecular fingerprints. To integrate ML with structure-based approaches, the T790M/L858R (TMLR) double-mutated epidermal growth factor receptor (EGFR) protein was chosen as it is one of the most important therapeutic drug targets in lung cancer.[13−15] The TMLR protein was then used for the virtual screening of approximately 400 000 NPs obtained from the COCONUT database.[16] The NPs having binding affinity better than the cocrystal ligand were subjected to screening using the developed ML model. Among the NPs that were predicted as actives by the ML model, the top three molecules having the lowest binding affinities were further evaluated for their stability using 100 ns molecular dynamics (MD) simulations. Overall, the integration of ML and a structure-based approach has led to the computational identification of indolocarbazole-based molecules that are not only strong and stable binders but are also predicted to have inhibitory activity against lung cancer cell lines.

METHOD

Phytomolecule Identification and Experimental Data Collection. The Naturally occurring Plant-based Anticancer Compound-Activity-Target (NPACT) database[17] is a repository of anticancerous phytomolecules with experimental in vitro activities recorded in the form of IC_{50}. Since the IC_{50} value of a compound denotes the minimum quantity of the compound that is required for the inhibition of a biological process by 50%, it is one of the important parameters of identifying the efficacy of a compound as an inhibitor. Therefore, NPACT was mined to extract phytomolecules with reported IC_{50} values against lung cancer cell lines. Additionally, to expand the collection of phytomolecules that have been studied against lung cancer at the in vitro level, literature mining of the PubMed database was performed using the R programming language. Papers that studied the role of phytomolecules in lung cancer but did not report any half-maximal inhibitory concentration (IC_{50}) were discarded. Finally, the compounds with IC_{50} < 10 μM were designated as actives, whereas compounds with IC_{50} > 10 μM were considered inactive. However, as bioactivity data have been obtained from different studies, some compounds were found to have multiple bioactivity values. Since the presence of variation in IC_{50} among duplicates introduces noise into the data set,[18] the majority rule was applied to categorize the molecule into the active−inactive data set. To decide the same, the number of entries with biological activity for a molecule was computed and if the bioactivity for the majority of the entries was <10 μM, then it was designated as active, while if the majority of the values were >10 μM, they were considered inactive. Also, molecules having an equal number of active and inactive values were removed from the data set. Additionally, the data set is devoid of inorganics and mixtures as only single compounds with bioactivity data have been collected from the literature. As a result, a data set of 649 phytomolecules was created.

Calculation of Molecular Fingerprints. Molecular fingerprints are a chemoinformatic representation of a molecular structure, wherein substructure key-based fingerprints convert molecular structures into bit strings based on the presence or absence of specific substructures. Two substructure key-based fingerprints, viz, Molecular ACCess Systems keys fingerprint (MACCS)[19] and Morgan2[20] containing 166-bit molecular descriptors and 1024-bit keysets, respectively, were calculated using the RDKit library in Python.

Model Building and Evaluation. To build the model, the data set was divided into training and external test validation based on an 80:20 ratio, respectively. The training/test split was accomplished using the train_test_split module from scikit-learn. The 80/20 rule of data splitting was implemented as it provides an adequate amount of molecules for training the models while keeping a sufficient amount of molecules for evaluating the trained models. The training set is used to construct the model by subjecting it to internal 5-fold cross-validation, whereas the validation data set is a blind set that is only used for assessing performance.

Four algorithms, viz., *k*-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGB), have been used in the current study for developing the ML models in scikit-learn. For all of the algorithms, a grid-search approach was used with recall as the objective function utilized for hyperparameter optimization. To evaluate the performance of the models, two validation techniques have been implemented. The training set is first subjected to 5-fold internal cross-validation, which is one of the most widely accepted methods. Herein, the training data is randomly split into five subsets and thereafter four subsets are used for training the model and the fifth subset is used for testing it. This process is repeated a total of five times such that each time a different subset is used as the test set and the average of the five outcomes yields the final result. One of the key features of this approach is that it utilizes all of the molecules for both training and testing. The model identified as best through 5-fold cross-validation was again subjected to evaluation using the external test validation set. Since this data set was not used for developing the model; therefore, using this data set provides confidence about the predictive capability of the model that has been developed. Further, for determining the quality of the models developed using each algorithm, statistical parameters like accuracy, specificity, sensitivity, and Matthews's correlation coefficient (MCC) were assessed. The receiver operating characteristic (ROC) curve was plotted, and its corresponding AUC was also calculated. The formulas used for calculating these parameters are given below

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

$$\text{specificity} = \frac{TN}{TN + FP}$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{MCC} = \frac{TP^{*}TN - FP^{*}FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP = true positive, FP = false positive, TN = true negative, and FN = false negative.

Y-scrambling Test. The Y-scrambling test was used to determine how likely it was that the top-performing model was selected by coincidence. To assess this, a new training set is

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

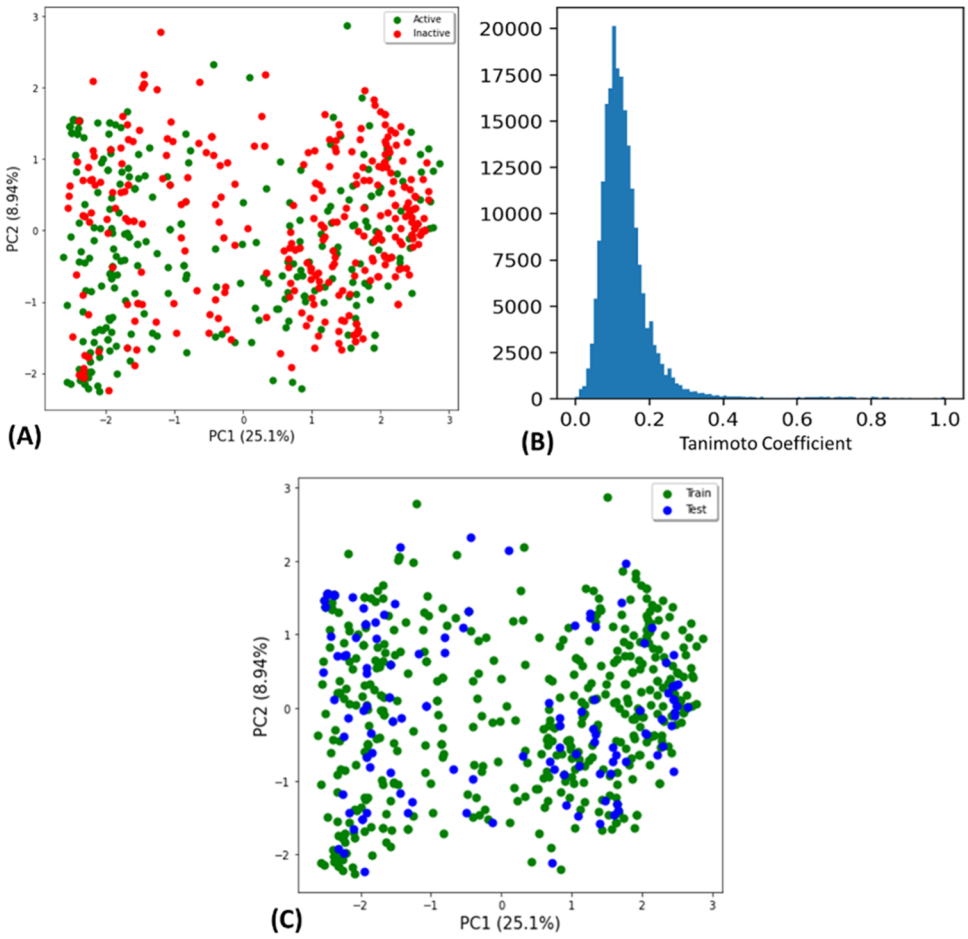


Figure 1. (A) Chemical distribution of 649 phytomolecules, where red dots represent actives and green dots represent inactive molecules. (B) Tanimoto similarity distribution. (C) Chemical space distribution of training and external test validation compounds.

produced by leaving the feature data set unchanged while arbitrarily rearranging the bioactivity labels. Using this scrambled training data set, a model with the same parameters as the best model was then constructed. Subsequently, the model's performance against the scrambled data set was examined using the external test validation set. This approach of scrambling and subsequent testing was repeated 30 times. If the performance of the scrambled models is lower than the best-performing model, then it provides confidence that the model generated is not by chance.

Library Preparation and ADMET Calculation. ~400 000 NPs were retrieved from the COCONUT database, which was neutralized, desalted, and minimized using the LigPrep module of Schrodinger. Thereafter, the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of the NPs were calculated using QikProp to identify molecules with structural features within the acceptable range.[21,22] A total of 24 pharmaceutically relevant properties and descriptors that determine the suitability of the molecule as a potential lead molecule were calculated. The properties are molecular weight, number of rotatable bonds, number of hydrogen bond donors and acceptors, dipole moment, ionization potential, electron affinity, various components of solvent accessible surface area (SASA, FOSA, FISA, PISA, WPSA), van der Waals surface area, total solvent accessible volume, globularity, polarizability, multiple measures of the partition coefficient (QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w), aqueous solubility, human serum albumin binding, blood/brain barrier partition coefficient, and metabolic descriptors. QikProp identified ~180 000 NPs that showed no violation of the above properties and were within the 95% range of known drugs (i.e., 0 stars). Therefore, ~180 000 NPs were considered suitable for virtual screening.

Virtual Screening Using Molecular Docking. The Protein Data Bank (PDB) was used to get the 3D crystal structure of the double-mutated EGFR protein harboring a reversible cocrystallized ligand (PDB ID: 5CAO).[23] 5CAO was selected for the current study as it has been shown to identify true actives early during virtual screening.[24] The protein preparation wizard of Schrodinger was used to preprocess, minimize, and refine the protein structure. The preprocessing step involved removing crystallographic waters, assigning bond orders, and adding missing hydrogens. To identify potential tight binders, a three-step docking approach was implemented, wherein the ADMET adherent molecules were first screened through high-throughput virtual screening (HTVS), then standard precision docking (SP), and finally, extra precision docking (XP). The cocrystal ligand of 5CAO was also redocked and the docking score was used as a control to identify high-affinity molecules.

Virtual Screening Using the ML Model. The molecules with binding affinities better than the cocrystal ligand were further screened using the ML model developed in this study to determine the antilung cancer potential of the molecules.

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

The NPs predicted to have antilung cancer activity by the binary classification model were taken for chemical structure classification using ClassyFire.[25]

Molecular Dynamics. The top three molecules with the lowest binding affinities were evaluated by using MD simulations. The three docked complexes were used as the starting conformation for the 100 ns MD simulations using Desmond.[26] The complexes were put into orthorhombic boxes of dimensions 10.0 Å × 10.0 Å × 10.0 Å, wherein the TIP3P water model and 0.15 M NaCl ions were used in the solution. Molecular minimization and molecular dynamics simulations were performed using the OPLS4 force field with default parameters. After equilibration, the three complexes were used for 100 ns MD simulations in the NPT ensemble. Upon completion of the simulation, root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), protein−ligand interactions, and interaction percentages throughout the MD simulation were used to assess the stability of the protein and ligands in the bound form.[27] Further, the relative free energy of ligand binding was estimated using molecular mechanics with generalized Bonn surface area (MM-GBSA) as given below

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - [\Delta G_{\text{protein}} + \Delta G_{\text{ligand}}]$$

wherein $\Delta G_{\text{binding}}$ represents the free energy of the protein−ligand complex, $\Delta G_{\text{protein}}$ and ΔG_{ligand} are energy values of the protein and ligand, respectively. Prime was used to calculate MM-GBSA by considering the simulation trajectories for the last 10 ns.

RESULTS AND DISCUSSION

Data Set Generation and Analysis. A data set of 649 phytomolecules having in vitro inhibitory activity against lung cancer cell lines was collected from the NPACT database and literature. Of these, 320 phytomolecules were designated as actives, while the rest 329 phytomolecules were labeled as inactives based on their biological activity. The chemical and structural diversity of the data set influences the process of model development; therefore, it was analyzed. The chemical space coverage of the data set was visualized by performing principal component analysis (PCA) using 166 MACCS descriptors (Figure 1A). The plot indicates that there is considerable overlap between active and inactive compounds as well as that they are scattered widely across chemical space, thereby reflecting significant chemical diversity. Also, to understand the structural diversity of the data set, the Tanimoto coefficient (T_c), which measures pairwise similarity for each pair of compounds, was calculated using Morgan2 fingerprints (Figure 1B). The distribution curve of T_c values shows that most pairs of compounds (∼90%) exhibit a similarity of less than 0.3, which indicates that the data set is diverse. Further, the chemical space of the training set (519 phytomolecules) and external test validation (130 phytomolecules) generated by the random stratified splitting method were examined through PCA (Figure 1C). It is clear from the plot that the external test validation overlapped with the distribution of the training set, revealing reasonable data set splitting. The above characteristics of the data sets revealed that it is suitable for model development and validation.

Table 1. Performance of ML Models for Fivefold Cross-Validation and Test Set[a]

	5-fold cross-validation				
Model	Acc (%)	AUC	S_P (%)	S_E (%)	MCC
RF_MACCS	76.5	0.81	77.6	75.4	0.53
RF_MORGAN2	73.2	0.80	76.4	69.9	0.47
SVM_MACCS	75.5	0.80	76.4	74.6	0.51
SVM_MORGAN2	74.5	0.81	77.5	71.5	0.49
KNN_MACCS	75.1	0.79	79.4	70.7	0.51
KNN_MORGAN2	70.7	0.71	77.9	63.3	0.42
XGB_MACCS	74.6	0.80	76.0	73.0	0.50
XGB_MORGAN2	72.1	0.78	73.8	70.3	0.44
	external test validation				
model	Acc (%)	AUC	S_P (%)	S_E (%)	MCC
RF_MACCS	73.1	0.80	75.0	71.2	0.46
RF_MORGAN2	66.9	0.80	75.0	59.1	0.35
SVM_MACCS	74.6	0.78	67.2	81.8	0.50
SVM_MORGAN2	70.8	0.79	76.6	65.2	0.42
KNN_MACCS	72.3	0.79	71.9	72.7	0.45
KNN_MORGAN2	71.5	0.71	71.9	71.2	0.43
XGB_MACCS	70.8	0.77	65.6	75.8	0.42
XGB_MORGAN2	72.3	0.80	75.0	69.7	0.45

[a]Legend—Acc: accuracy; SP: specificity; SE: sensitivity.

Model Development and Evaluation. In the current study, both MACCS and Morgan2 fingerprints along with four different ML algorithms (i.e., KNN, SVM, RF, and XGBoost) were used to build eight types of models, which were evaluated using 5-fold cross-validation and independent external test validation. It is observed that the models developed with MACCS fingerprints showed better sensitivity than Morgan2 for each of the ML algorithms. Among the eight models, the 5-fold cross-validated outcome shows that the model built with MACCS fingerprint and RF classifier (i.e., RF_MACCS model) exhibited the highest sensitivity and accuracy and therefore was chosen as the best-performing model (Table 1). It showed 75.4% sensitivity, 77.6% specificity, and 76.5% accuracy with a 0.53 MCC on the training set. On the test set, the model shows 75% specificity, 71.2% sensitivity, and 73.1% accuracy with 0.46 MCC (Table 1). Finally, the RF model generated using the MACSS fingerprint has been selected, as it shows the best performance.

RF_MACCS Model Validation Using Y-randomization. Apart from 5-fold internal cross-validation and test set validation, the best-performing model, i.e., RF_MACCS, was also validated using Y-randomization to check the presence of coincidental correlations. Herein, the average accuracy, sensitivity, specificity, MCC, and AUC for the model after 30 randomized shuffles were approximately 51%, 49%, 53%, 0.02, and 0.50. As the shuffled models exhibited poor performance in comparison to the selected model, it thus indicates that the RF_MACCS model is statistically robust.

k-Means Clustering of Lung Cancer Inhibitors. The structural features of the 649 antilung cancer molecules were investigated by the *k*-means clustering approach using the 166 MACCS fingerprints in scikit-learn. In order to cluster the data, the *k*-means approach divides the molecules into n number of groups having equal variance and minimizes a criterion called the inertia within-cluster sum-of-squares. Further, t-SNE[28] was used to reduce the dimensionality of the data and visualize the eight clusters in two dimensions (Figure 2). Thereafter, the number of actives and inactives present in each cluster was calculated to identify whether a cluster contains actives or inactives in the majority. Clusters 1,

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

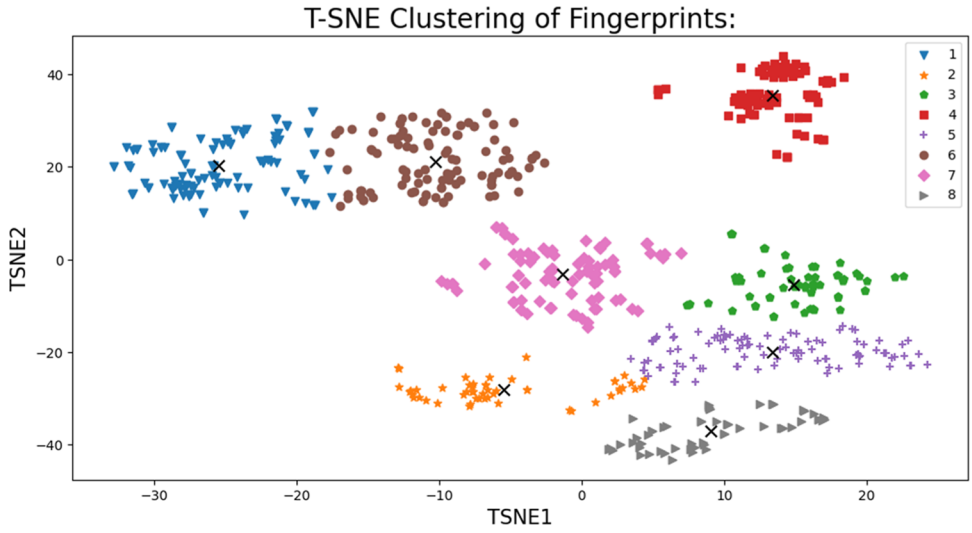


Figure 2. t-SNE representation of the eight clusters created from 649 inhibitors using *k*-means clustering. TSNE1 and TSNE2 represent the two dimensions reduced from the 166 features of MACCS fingerprint and the black cross represents the centroid of each cluster.

Table 2. Characteristics of the Top 10 MACCS Fingerprints[a]

sl no.	MACCS keys	description	MDI	FS
1	bit 96	5 M ring	0.0278	30.39
2	bit 136	O═A > 1	0.0270	16.00
3	bit 128	ACH2AAACH2A	0.0265	33.09
4	bit 129	ACH2AACH2A	0.0264	32.74
5	bit 53	QHAAAQH	0.0243	−18.26
6	bit 72	OAAO	0.0240	22.20
7	bit 83	QAAAA@1	0.0240	29.12
8	bit 104	QHACH2A	0.0215	24.70
9	bit 154	C═O	0.0210	7.22
10	bit 90	QHAACH2A	0.0200	19.49

[a]MDI: Mean decrease in impurity. FS: Difference in the frequency of a fingerprint in the active and inactive molecules. Atom symbols: A = any valid periodic table element, Q = hetero atoms, O = oxygen, C = carbon, H = hydrogen. Bond symbols: = means double bond.

4, and 8 were found to be dominated by active molecules, wherein the percentage of active molecules ranged from 60 to 90%. Clusters 2, 3, and 5 were majorly represented with inactive molecules, and the percentage of inactive molecules was around 70%. While clusters 6 and 7 had comparable numbers of actives and inactives. To further understand the chemical scaffolds present in these clusters, the molecules were classified using ClassyFire.

Cluster 1 consists of 100 molecules, among which 63% were actives and 37% were inactives. The active molecules were mainly prenol lipids and steroid derivatives like triterpenoids, quassinoids, triterpene saponins, limonoids, withanolide derivatives, cardenolide glycosides, and steroidal saponins. Cluster 4 consisted of 77 molecules, among which 91% were composed of actives. This cluster contains the largest number of actives and belongs to the annonaceous acetogenin category.[29] In cluster 8, a total of 55 molecules were present, among which 60% were active and 40% were inactive. Among the active molecules, alkaloids were the major class present,

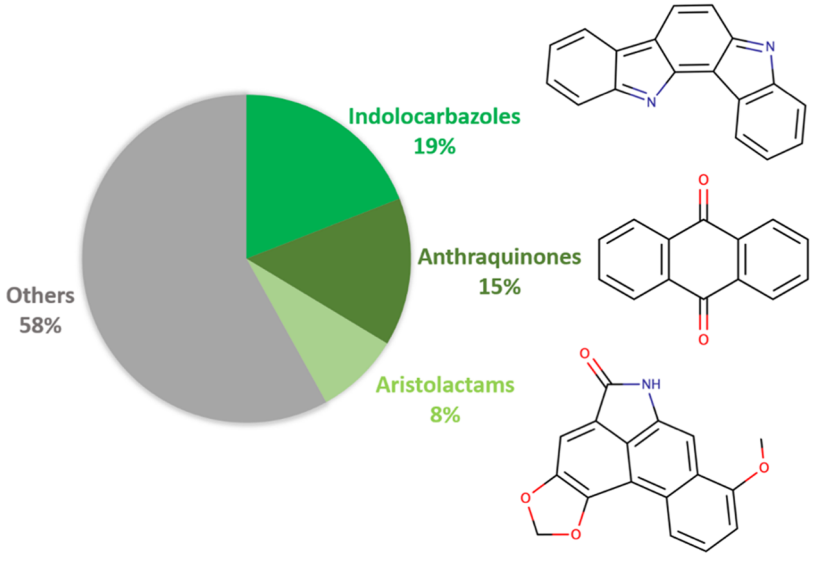


Figure 3. Top three classes of molecules identified through integrated molecular docking and the ML-based approach.

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

Table 3. Binding Affinity and Free Energies of the Protein−Ligand Complexes Calculated through Molecular Docking and MM-GBSA (All Values Reported in kcal/mol)

ligand	binding affinity	binding free energy (ΔG_{bind})	Coulomb energy ($\Delta E_{coulomb}$)	lipophilic energy (ΔE_{lipo})	van der Waals energy (ΔE_{vdw})
cocrystal	−10.15	−64.77 ± 2.89	−12.86 ± 1.74	−12.99 ± 0.8	−57.89 ± 2.31
S5—3509	−13.83	−73.41 ± 3.84	−34.83 ± 4.92	−16.66 ± 0.89	−57.57 ± 2.38
S6—20 862	−13.50	−67.13 ± 3.67	−28.1 ± 5.75	−15.27 ± 0.71	−52.5 ± 2.28
S7—9517	−13.24	−61.76 ± 3.88	−28.12 ± 5.02	−14.62 ± 0.88	−51.61 ± 2.52

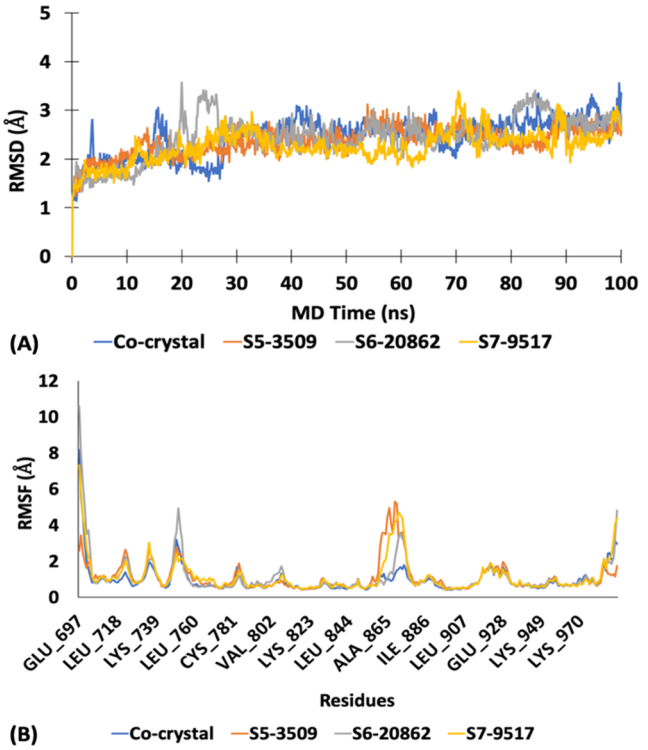


Figure 4. (A) Protein backbone RMSD in reference to the first frame of the MD simulation. (B) RMSF of the Cα atoms in the protein backbone.

wherein subcategories like aporphines and oxo-aporphines were found. Moreover, in this cluster, several other families of compounds were also observed, such as benzoquinolines, pyranoquinolines, benzofurans, and indoles.

In cluster 2, 52 molecules were present, among which 71.2% of the molecules were inactive, while the rest were active. These inactive molecules belonged to the phenylpropanoids and polyketide class of natural products like flavonoids, isoflavonoids, cinnamic acids, and their derivatives as well as furanoid lignans. Cluster 3 contained 60 molecules, of which 70% were inactive and majorly belonged to the phenylpropanoids and polyketide classes similar to the inactives of cluster 2. However, the major subcategories in this cluster were linear 1,3-diarylpropanoids and stilbenes. Apart from these, a few molecules were of the benzenoid family as well. Among the 109 molecules present in cluster 5, 71.6% of molecules were inactive. These molecules were also found to belong to the phenylpropanoids and polyketide class of natural products like the inactives of clusters 2 and 3 as well as furanoid lignans of cluster 2.

In cluster 6, there were 107 molecules, among which 53 were active and 54 were inactive. Prenol lipids like terpenoids were heavily present in both the active and inactive groups along with steroids and their derivatives. In the case of cluster 7, 58% of the molecules were inactive, whereas the remaining 42% were active. In both data sets, phenylpropanoids and polyketides were the major categories along with benzopyran-based compounds.

Overall, it was observed that prenol lipids, steroid derivatives, withanolide derivatives, and annonaceous acetogenins are the major classes of compounds that show inhibitory activity against lung cancer (actives), while most of the inactive molecules are from the phenylpropanoids and polyketide classes of molecules.

Feature Importance and Analysis of MACCS Fingerprints. Feature importance is an important part of ML model building, as it is capable of identifying the contribution of each feature or fingerprint to the model's prediction. Using mean decrease in impurity (MDI) or Gini importance, the top 10 important fingerprints with respect to the RF_MACCS model were calculated (Table 2). This approach evaluates the capability of each feature to reduce uncertainty. MDI was calculated using 'feature_importances' in scikit-learn. For these 10 important fingerprints, the difference in frequency in the active and inactive data sets was also computed. For each MACCS fingerprint, the frequency difference has been calculated as given below and the ten most important MACCS keys as well as their characteristics are given in Table 2.

$$F_i^{\mathrm{A}} = \frac{\sum_{j=1}^{\mathrm{NA}} D_i^j}{\mathrm{NA}} \times 100 \quad (1)$$

$$F_i^{\mathrm{I}} = \frac{\sum_{j=1}^{\mathrm{NI}} D_i^j}{\mathrm{NI}} \times 100 \quad (2)$$

where F_i^{A} and F_i^{I} are the mean of the ith fingerprint in active (A) and inactive (I) molecules, respectively; NA and NI are the total number of molecules present in the active and inactive data sets; and D_i^j is the value of the ith fingerprint for jth molecule (either 0 or 1). Using eqs 1 and 2, the fingerprint score (FS) is then calculated for each fingerprint as follows

$$\mathrm{FS}_i = F_i^{\mathrm{A}} - F_i^{\mathrm{I}} \quad (3)$$

where FS_i is the difference in frequency of the ith fingerprint.

ADMET Screening of the COCONUT Database. As NP molecules are often complex molecules, it becomes difficult for them to get translated into drugs, and therefore, the COCONUT natural product database, which contains approximately 400 000 NPs, was screened for their pharmacokinetic profile using QikProp. QikProp generates molecular descriptors from the chemical structure of phytomolecules and utilizes them to predict the ADMET properties. It calculates properties that fall beyond the 95% range of known drugs and designates a star if the value lies outside the accepted threshold. As a result, approximately 180 000 NPs that exhibited no outliers in their ADMET properties, i.e., those

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

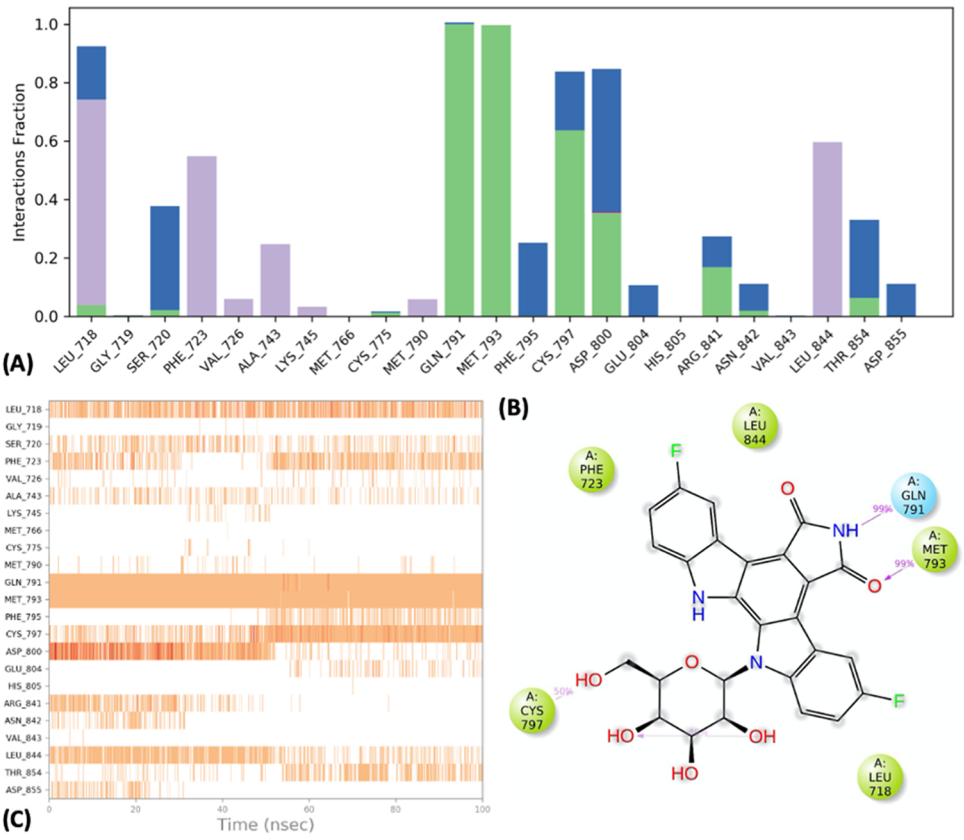


Figure 5. Interactions of the S5—3509 bound complex with (A) the various residues of the protein structure and the different types of interaction they are involved in. (B) Ligand and surrounding residues are shown in a 2D interaction diagram along with their percentage. (C) H-bonds formed during 100 ns simulation time.

molecules having "0" stars were shortlisted as they have a higher probability of showing ADMET adherent behavior.

Molecular Docking. Since the double-mutated EGFR is one of the primary oncogenic proteins involved in the etiology of a large fraction of lung cancer cases, this protein was selected for molecular docking to identify molecules capable of binding to the protein. The crystal structure of the double-mutated EGFR protein was therefore subjected to a three-step docking approach consisting of HTVS, SP, and XP. These different docking modes are designed, such that the level of precision increases consecutively from HTVS to SP to XP. HTVS is useful to screen large chemical libraries due to its simplified scoring function and faster optimization algorithm, which enables the identification of potential hits in a relatively shorter time frame. Therefore, initially, the 180 000 ADMET adherent NPs were screened through HTVS, which led to the identification of the top 60 000 molecules with the lowest binding affinities. As SP mode offers a more accurate scoring function compared to HTVS, the 60 000 molecules were further screened using SP to identify the top 6000 molecules with the lowest binding affinity. The XP mode provides the highest precision docking in Glide and focuses on the accurate prediction of ligand binding modes and binding affinities. Because it employs advanced sampling techniques and a more sophisticated scoring function, it captures subtle interactions between ligands and proteins compared to HTVS and SP. Finally, the top 6000 molecules identified from SP were screened using XP mode along with the cocrystal ligand. The cocrystal ligand was used as a positive control, and its docking score which resembles the binding affinity of the ligand toward the protein active site was then used as a reference for identifying better binders. The cocrystal ligand was found to dock with a score of −10.15 kcal/mol, and therefore, any molecule that scored with a more negative docking score (i.e., better binding) was selected. As a result, 616 NP molecules which show docking scores ranging from −10.16 to −13.83 kcal/mol were selected for scaffold analysis as they exhibited higher binding affinity than that of the cocrystal ligand.

ML Model-Based Screening. The anticancer potential of the 616 NPs was further evaluated using the RF_MACCS ML model to predict their activity against lung cancer. Of these, 205 molecules were predicted as active, which were majorly found to harbor indolocarbazole (19%), anthraquinone (15%), and aristolactam (8%) scaffolds (Figure 3). The anticancer activity of indolocarbazoles has been studied against a wide range of cellular targets and has given rise to potent drugs like staurosporine, midostaurin, etc.[30,31] Also, anthraquinone and aristolactam classes of molecules target various key proteins in cancer cells, which has resulted in the development of inhibitors like doxorubicin, epirubicin, and aristolactam IIIa.[32,33] It was also observed that the three top-scoring molecules with the most favorable binding affinity exhibited docking scores between −13.24 and −13.83 kcal/mol and all of them had indolocarbazole backbone (Table 3). Therefore, the top 3 molecules were selected for molecular dynamics analysis.

Molecular Dynamics to Ascertain the Stability of the Protein−Ligand Complexes. The stability and dynamic behavior of the top three NPs with the lowest binding affinities, i.e., compounds S5—3509 (−13.83 kcal/mol), S6—

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

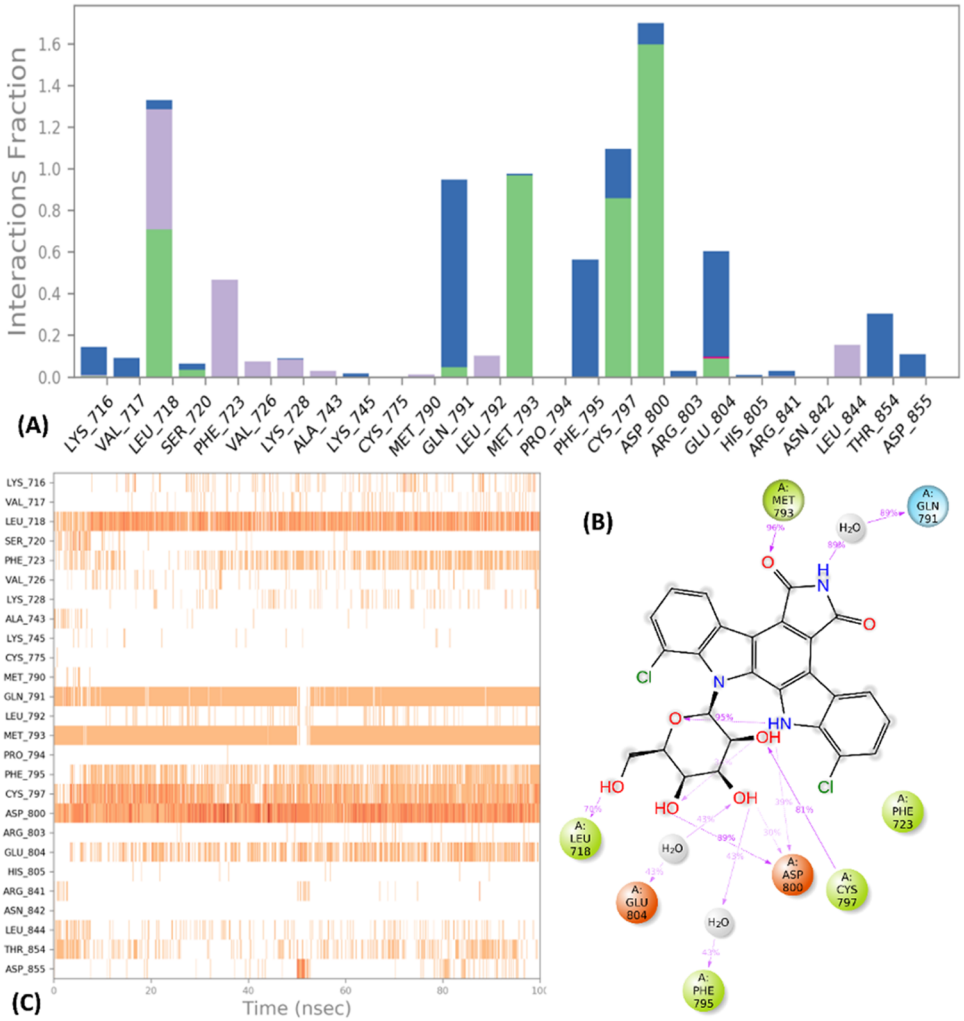


Figure 6. Interactions of the S6—20 862 bound complex with (A) the various residues of the protein structure and the different types of interaction they are involved in. (B) Ligand and surrounding residues are shown in a 2D interaction diagram along with their percentage. (C) H-bonds formed during 100 ns simulation time.

20862 (−13.50 kcal/mol), and S7—9517 (−13.24 kcal/mol), were thereafter checked using 100 ns MD simulations. The cocrystal ligand of 5CAO was also subjected to a 100 ns simulation as a control. The RMSD analysis of the protein backbone indicated that apart from minor fluctuations, all of the systems stabilized after 30 ns of simulation time (Figure 4A). The average RMSD values of S7—9517, S5—3509, and S6—20862 were 2.26, 2.36, and 2.45 Å, respectively, which were either lower or similar to that of the cocrystal ligand (2.45 Å) (Figure 4A). This demonstrates that the identified potential inhibitors are stable within the active site of the protein. Also, to test the flexibility of residues that may have contributed to overall movements in the system, the RMSF of the Cα chain was calculated for the protein−ligand complexes (Figure 4B). The RMSF plot shows that the behavior of the residues is almost similar in all of the complexes, thereby demonstrating no significant perturbation in the protein due to the binding of the identified ligands.

Molecular Dynamics-Based Interaction Analysis. The interaction study of the ligand−receptor complexes for a 100 ns simulation period indicated that all of the three lead molecules S5—3509, S6—20862, and S7—9517 established hydrogen bonds with the protein backbone via Gln791 and Met793, similar to the cocrystal ligand (Figures 5−7 and S1). The S6—20862 molecule had maintained the H-bond with Met793 for 96% of the total simulation time like the cocrystal ligand, while in the case of both S5—3509 and S7—9517, this interaction was maintained for 99% of the MD time. On the other hand, the H-bond with Gln791 was present for 91% of the simulation time in the cocrystal ligand-bound complex, whereas in the case of S5—3509, it was present for 99% of the time. This interaction was also present in the S6—20862 and S7—9517 bound proteins; however, it was water-mediated and maintained for 90 and 89% of the time, respectively. Additionally, these three ligands showed H-bond interaction with Cys797 for different time intervals, *viz*, 50% in S5—3509, 81% in S6—20 862, and 91% in S7—9517. Moreover, the ligand S6—20 862 demonstrated additional H-bonds with Asp800 and Leu718, while S7—9517 formed H-bond with Asp800. Hydrophobic interactions, in addition to hydrogen bonds, are also crucial for the binding of drugs to their targets. The cocrystal ligand displayed hydrophobic interactions with Leu718 and Phe723, which have been maintained in all three ligand-bound complexes (Figure 5−7 and S1).

Binding Affinity Calculations and Per-Residue-Decomposition Analysis. The MMBSA method was used to

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

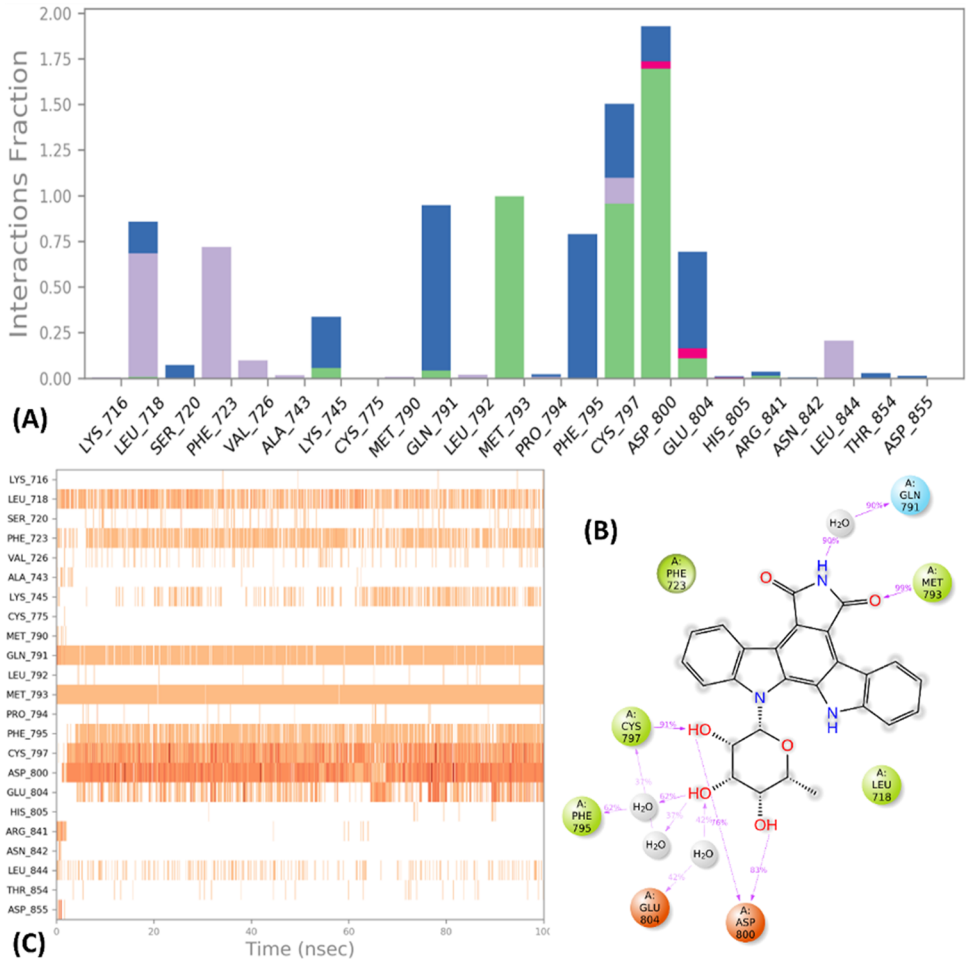


Figure 7. Interactions of the S7—9517 bound complex with (A) the various residues of the protein structure and the different types of interaction they are involved in. (B) Ligand surrounding residues are shown in a 2D interaction diagram along with their percentage; (C) H-bonds formed during 100 ns simulation time.

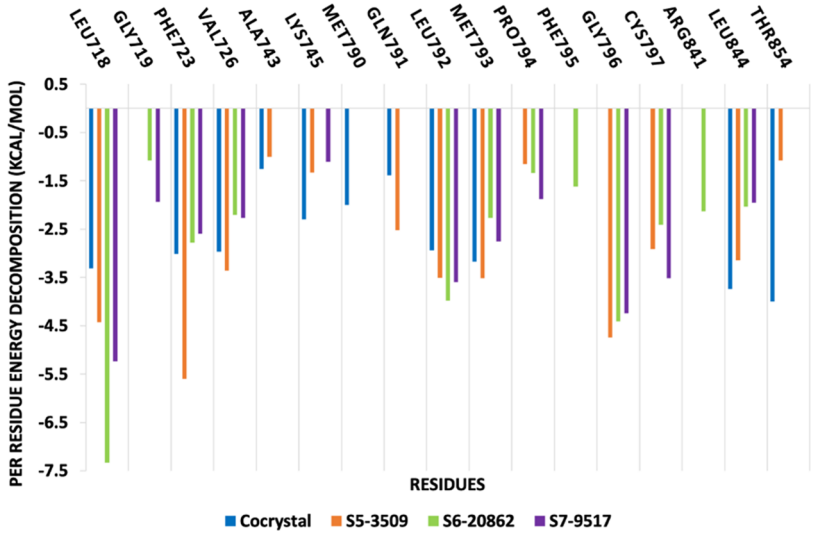


Figure 8. Per residue energy decomposition analysis of the identified inhibitors against the EGFR mutant protein.

evaluate the ligand's binding affinity and stability using the snapshots of the protein−ligand binding over the last 10 ns of the simulation. It was found that binding free energies (ΔG_{bind}) of S5—3509 (−73.41 ± 3.84 kcal/mol) and S6—20 862 (−67.13 ± 3.67 kcal/mol) were better than the cocrystal ligand (−64.77 ± 2.89), while for S7—9517, the value was comparable (−61.76 ± 3.88 kcal/mol) (Table 3). The contribution of the individual energies in the overall binding energy was also calculated and it was observed that the Coulomb energy ($\Delta E_{coulomb}$) and lipophilic energy (ΔE_{lipo}) of

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

the three identified ligands were better than the cocrystal ligand, while the van der Waals energy (ΔE_{vdw}) was either comparable or lower than the cocrystal ligand (Table 3).

Additionally, the contribution of the individual amino acids in the protein−ligand complex was also examined, wherein the residues having ΔG_{bind} less than −1 kcal/mol were considered to be the key contributors to stability. The contribution of the residues Leu718, Phe723, Leu792, Met793, and Leu844 was found to be present in the three potential inhibitors and the cocrystal ligand (Figure 8). However, the contributions of a few amino acids, such as Pro794, Phe795, Gly796, Cys797, and Arg841, were observed only in the identified molecules. Moreover, Leu718, Phe723, and Gly796 were observed to have remarkable contributions to ligand stability with the ΔG_{bind} values ranging from approximately −3 to −7.3 kcal/mol. Overall, the three identified NPs having indolocarbazole backbone show conserved hydrogen and hydrophobic interactions and exhibit binding energies better than the cocrystal complex.

Structural Similarity of Known Drugs with the Computationally Identified Potential Inhibitors. The three compounds identified from the current study were checked for their structural similarity with known drugs using the DrugBank database for a better understanding of the structure−activity relationship of these molecules. Using a structural similarity cutoff of >80%, seven drugs—rebeccamycin, becatecarin, edotecarin, lestaurtinib, staurosporine, K-252a, and 7-hydroxystaurosporine—were identified. Thereafter, a literature survey was performed to find the bioactivity of these molecules against EGFR and its mutants. Herein, we identified staurosporine and its derivative 7-hydroxystaurosporine as mutant-selective EGFR inhibitors. Zhao et al. have shown that staurosporine inhibits T790M/L858R mutation and wild-type EGFR with IC_{50} values of 1.1 and 74.2 nM, respectively.[34] Similarly, Mansour et al. showed that the IC_{50} inhibitory value of staurosporine against T790M/L858R mutation was 2.3 nM in comparison to 125 nM for wild-type EGFR.[35] Overall, several studies have documented the selectivity of staurosporine against the T790M/L858R mutation of EGFR over wild-type EGFR ranging from 55 to ∼300 folds.[35−39] Moreover, staurosporine had also demonstrated dose-dependent inhibition of cell proliferation of A549 cells with IC_{50} values ranging from 4.29 to 9.50 μM,[40,41] respectively. In addition, staurosporine also inhibits other lung cancer cell lines, such as N417, H209, and Ma-31, with IC_{50} values of 54, 29, and 602 nM,[42] respectively, demonstrating its potency. The similarity between the compounds identified in the current study and the drug staurosporine is primarily due to the indolocarbazole scaffold that is common in all of these compounds. Thus, in the current study, indolocarbazole scaffold-based molecules have been computationally identified as the most favorable, which show mutant selective activity against EGFR.

CONCLUSIONS

One of the primary hurdles of anticancer therapeutics is the development of resistance that renders the known drugs ineffective, a major issue in lung cancer as well. Therefore, there is a continuous need to identify molecules that can be developed as potential leads to lung cancer. As natural products are an excellent source of NCEs that have proven biological activities due to their structural features, we have utilized the knowledge of antilung cancer phytomolecules to develop ML models. Various ML algorithms and chemical fingerprints were used to develop multiple models, among which the random forest model having MACCS fingerprint showed the best performance. To validate the model capability, an external test validation set and Y-randomization technique were used, which indicated satisfactory performance. To analyze the feature importance, mean decrease in impurity was calculated and k-means clustering with t-SNE was performed for chemical space exploration. The model was then used to screen NPs having binding affinity better than the cocrystal ligand of the EGFR double-mutant protein. Further, analysis revealed that compounds with an indolocarbazole scaffold were majorly present in top-scoring docked molecules, and therefore, the top three protein−ligand complexes were evaluated for their stability in the protein active site using MD simulations and MM-GBSA analyses. Thus, the integration of ML and a structure-based approach allowed us to identify indolocarbazole-based molecules that exhibit inhibitory potential against lung cancer.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c07338.

Results of molecular dynamics of the cocrystal ligand of 5CAO (PDF)

AUTHOR INFORMATION

Corresponding Author

Subhash M. Agarwal − *Bioinformatics Division, ICMR-National Institute of Cancer Prevention and Research, Noida 201301 Uttar Pradesh, India;* orcid.org/0000-0001-6685-789X; Email: smagarwal@yahoo.com

Authors

Agneesh Pratim Das − *Bioinformatics Division, ICMR-National Institute of Cancer Prevention and Research, Noida 201301 Uttar Pradesh, India; Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida 201313 Uttar Pradesh, India*

Puniti Mathur − *Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida 201313 Uttar Pradesh, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c07338

Author Contributions

Conceptualization: S.M.A.; methodology: S.M.A., A.P.D.; formal analysis and investigation: S.M.A., A.P.D; writing—original draft preparation: A.P.D.; writing—review and editing: S.M.A., P.M.; funding acquisition: S.M.A.; resources: S.M.A.; supervision: S.M.A., P.M.

Funding

SMA acknowledges the support of the Indian Council of Medical Research (Grant No. 59/03/2019/Online/BMS/TRM).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

SMA acknowledges The Director, NICPR, for institutional support.

REFERENCES

(1) Siegel, R. L.; Miller, K. D.; Wagle, N. S.; Jemal, A. Cancer Statistics, 2023. *CA Cancer J. Clin.* **2023**, *73* (1), 17−48.

(2) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71* (3), 209−249.

(3) Yuan, M.; Huang, L.-L.; Chen, J.-H.; Wu, J.; Xu, Q. The Emerging Treatment Landscape of Targeted Therapy in Non-Small-Cell Lung Cancer. *Signal Transduction Targeted Ther.* **2019**, *4*, 61.

(4) Atanasov, A. G.; Waltenberger, B.; Pferschy-Wenzig, E.-M.; Linder, T.; Wawrosch, C.; Uhrin, P.; Temml, V.; Wang, L.; Schwaiger, S.; Heiss, E. H.; Rollinger, J. M.; Schuster, D.; Breuss, J. M.; Bochkov, V.; Mihovilovic, M. D.; Kopp, B.; Bauer, R.; Dirsch, V. M.; Stuppner, H. Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review. *Biotechnol. Adv.* **2015**, *33* (8), 1582−1614.

(5) Das, A. P.; Agarwal, S. M. Recent Advances in the Area of Plant-Based Anti-Cancer Drug Discovery Using Computational Approaches. *Mol. Diversity* **2023**, 1−25.

(6) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770−803.

(7) Yadav, I. S.; Nandekar, P. P.; Srivastavaa, S.; Sangamwar, A.; Chaudhury, A.; Agarwal, S. M. Ensemble Docking and Molecular Dynamics Identify Knoevenagel Curcumin Derivatives with Potent Anti-EGFR Activity. *Gene* **2014**, *539* (1), 82−90.

(8) Sadybekov, A. V.; Katritch, V. Computational Approaches Streamlining Drug Discovery. *Nature* **2023**, *616* (7958), 673−685.

(9) Greener, J. G.; Kandathil, S. M.; Moffat, L.; Jones, D. T. A Guide to Machine Learning for Biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (1), 40−55.

(10) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discovery* **2019**, *18* (6), 463−477.

(11) Singh, H.; Singh, S.; Singla, D.; Agarwal, S. M.; Raghava, G. P. QSAR Based Model for Discriminating EGFR Inhibitors and Non-Inhibitors Using Random Forest. *Biol. Direct* **2015**, *10*, No. 10.

(12) Dhiman, K.; Agarwal, S. M. NPred: QSAR Classification Model for Identifying Plant Based Naturally Occurring Anti-Cancerous Inhibitors. *RSC Adv.* **2016**, *6* (55), 49395−49400.

(13) Agarwal, S. M.; Pal, D.; Gupta, M.; Saini, R. Insight into Discovery of Next Generation Reversible TMLR Inhibitors Targeting EGFR Activating and Drug Resistant T790M Mutants. *Curr. Cancer Drug Targets* **2017**, *17* (7), 617−636.

(14) Saini, R.; Fatima, S.; Agarwal, S. M. TMLRpred: A Machine Learning Classification Model to Distinguish Reversible EGFR Double Mutant Inhibitors. *Chem. Biol. Drug Des.* **2020**, *96* (3), 921−930.

(15) Leonetti, A.; Sharma, S.; Minari, R.; Perego, P.; Giovannetti, E.; Tiseo, M. Resistance Mechanisms to Osimertinib in EGFR-Mutated Non-Small Cell Lung Cancer. *Br. J. Cancer* **2019**, *121* (9), 725−737.

(16) Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M. A.; Steinbeck, C. COCONUT Online: Collection of Open Natural Products Database. *J. Cheminform.* **2021**, *13* (1), 2.

(17) Mangal, M.; Sagar, P.; Singh, H.; Raghava, G. P. S.; Agarwal, S. M. NPACT: Naturally Occurring Plant-Based Anti-Cancer Compound-Activity-Target Database. *Nucleic Acids Res.* **2013**, *41* (D1), D1124−D1129.

(18) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model.* **2016**, *56* (7), 1243−1252.

(19) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273−1280.

(20) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5* (2), 107−113.

(21) Sharma, A.; Sharma, S.; Gupta, M.; Fatima, S.; Saini, R.; Agarwal, S. M. Pharmacokinetic Profiling of Anticancer Phytocompounds Using Computational Approach. *Phytochem. Anal.* **2018**, *29* (6), 559−568.

(22) Sharma, S.; Gupta, M.; Sharma, A.; Agarwal, S. M. Oral Bioavailability of Naturally Occurring Anticancer Phytomolecules. *Lett. Drug Des. Discovery* **2018**, *15* (11), 1180−1188.

(23) Heald, R.; Bowman, K. K.; Bryan, M. C.; Burdick, D.; Chan, B.; Chan, E.; Chen, Y.; Clausen, S.; Dominguez-Fernandez, B.; Eigenbrot, C.; Elliott, R.; Hanan, E. J.; Jackson, P.; Knight, J.; La, H.; Lainchbury, M.; Malek, S.; Mann, S.; Merchant, M.; Mortara, K.; Purkey, H.; Schaefer, G.; Schmidt, S.; Seward, E.; Sideris, S.; Shao, L.; Wang, S.; Yeap, K.; Yen, I.; Yu, C.; Heffron, T. P. Noncovalent Mutant Selective Epidermal Growth Factor Receptor Inhibitors: A Lead Optimization Case Study. *J. Med. Chem.* **2015**, *58* (22), 8877−8895.

(24) Das, A. P.; Nandekar, P.; Mathur, P.; Agarwal, S. M. A Systematic Pipeline of Protein Structure Selection for Computer-Aided Drug Discovery: A Case Study on T790M/L858R Mutant EGFR Structures. *Protein Sci.* **2023**, *32*, No. e4740.

(25) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminform.* **2016**, *8*, 61.

(26) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing; SC '06*; Association for Computing Machinery: New York, NY, USA, 2006; p 84 DOI: 10.1145/1188455.1188544.

(27) Nandhini, P.; Gupta, P. K.; Mahapatra, A. K.; Das, A. P.; Agarwal, S. M.; Mickymaray, S.; Alothaim, A. S.; Rajan, M. In-Silico Molecular Screening of Natural Compounds as a Potential Therapeutic Inhibitor for Methicillin-Resistant Staphylococcus Aureus Inhibition. *Chem.-Biol. Interact.* **2023**, *374*, No. 110383.

(28) van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9* (86), 2579−2605.

(29) Mangal, M.; Imran Khan, M.; Mohan Agarwal, S. Acetogenins as Potential Anticancer Agents. *Anti-Cancer Agents Med. Chem.* **2016**, *16* (2), 138−159.

(30) Wang, Y.; Yang, H.; Liu, H.; Huang, J.; Song, X. Effect of staurosporine on the Mobility and Invasiveness of Lung Adenocarcinoma A549 Cells: An in Vitro Study. *BMC Cancer* **2009**, *9*, 174.

(31) Stone, R. M.; Mandrekar, S. J.; Sanford, B. L.; Laumann, K.; Geyer, S.; Bloomfield, C. D.; Thiede, C.; Prior, T. W.; Döhner, K.; Marcucci, G.; Lo-Coco, F.; Klisovic, R. B.; Wei, A.; Sierra, J.; Sanz, M. A.; Brandwein, J. M.; de Witte, T.; Niederwieser, D.; Appelbaum, F. R.; Medeiros, B. C.; Tallman, M. S.; Krauter, J.; Schlenk, R. F.; Ganser, A.; Serve, H.; Ehninger, G.; Amadori, S.; Larson, R. A.; Döhner, H. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3Mutation. *N. Engl. J. Med.* **2017**, *377* (5), 454−464.

(32) Malik, M. S.; Alsantali, R. I.; Jassas, R. S.; Alsimaree, A. A.; Syed, R.; Alsharif, M. A.; Kalpana, K.; Morad, M.; Althagafi, I. I.; Ahmed, S. A. Journey of Anthraquinones as Anticancer Agents - a Systematic Review of Recent Literature. *RSC Adv.* **2021**, *11* (57), 35806−35827.

(33) Li, L.; Wang, X.; Chen, J.; Ding, H.; Zhang, Y.; Hu, T.-C.; Hu, L.-H.; Jiang, H.-L.; Shen, X. The Natural Product Aristolactam AIIIa as a New Ligand Targeting the Polo-Box Domain of Polo-like Kinase 1 Potently Inhibits Cancer Cell Proliferation. *Acta Pharmacol. Sin.* **2009**, *30* (10), 1443−1453.

(34) Zhao, J.; Fang, L.; Zhang, X.; Liang, Y.; Gou, S. Synthesis and Biological Evaluation of New [1,2,4]Triazolo[4,3-a]Pyridine Deriva-

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539

tives as Potential c-Met Inhibitors. *Bioorg. Med. Chem.* **2016**, *24* (16), 3483−3493.
(35) Mansour, T. S.; Pallepati, R. R.; Basetti, V. Potent Dual EGFR/Her4 Tyrosine Kinase Inhibitors Containing Novel (1,2-Dithiolan-4-Yl)Acetamides. *Bioorg. Med. Chem. Lett.* **2020**, *30* (16), No. 127288.
(36) Ai, X.; Sun, Y.; Wang, H.; Lu, S. A Systematic Profile of Clinical Inhibitors Responsive to EGFR Somatic Amino Acid Mutations in Lung Cancer: Implication for the Molecular Mechanism of Drug Resistance and Sensitivity. *Amino Acids* **2014**, *46* (7), 1635−1648.
(37) Günther, M.; Juchum, M.; Kelter, G.; Fiebig, H.; Laufer, S. Lung Cancer: EGFR Inhibitors with Low Nanomolar Activity against a Therapy-Resistant L858R/T790M/C797S Mutant. *Angew. Chem., Int. Ed.* **2016**, *55* (36), 10890−10894.
(38) Song, X.; Liu, X.; Ding, X. staurosporine Scaffold-Based Rational Discovery of the Wild-Type Sparing Reversible Inhibitors of EGFR T790M Gatekeeper Mutant in Lung Cancer with Analog-Sensitive Kinase Technology. *J. Mol. Recognit.* **2017**, *30* (4), No. e2590.
(39) Juchum, M.; Günther, M.; Döring, E.; Sievers-Engler, A.; Lämmerhofer, M.; Laufer, S. Trisubstituted Imidazoles with a Rigidized Hinge Binding Motif Act As Single Digit NM Inhibitors of Clinically Relevant EGFR L858R/T790M and L858R/T790M/C797S Mutants: An Example of Target Hopping. *J. Med. Chem.* **2017**, *60* (11), 4636−4656.
(40) Al-Warhi, T.; El Kerdawy, A. M.; Said, M. A.; Albohy, A.; Elsayed, Z. M.; Aljaeed, N.; Elkaeed, E. B.; Eldehna, W. M.; Abdel-Aziz, H. A.; Abdelmoaz, M. A. Novel 2-(5-Aryl-4,5-Dihydropyrazol-1-Yl)Thiazol-4-One as EGFR Inhibitors: Synthesis, Biological Assessment and Molecular Docking Insights. *Drug Des., Dev. Ther.* **2022**, 1457−1471.
(41) Ahmed, E. Y.; Elserwy, W. S.; El-Mansy, M. F.; Serry, A. M.; Salem, A. M.; Abdou, A. M.; Abdelrahman, B. A.; Elsayed, K. H.; Abd Elaziz, M. R. Angiokinase Inhibition of VEGFR-2, PDGFR and FGFR and Cell Growth Inhibition in Lung Cancer: Design, Synthesis, Biological Evaluation and Molecular Docking of Novel Azaheterocyclic Coumarin Derivatives. *Bioorg. Med. Chem. Lett.* **2021**, *48*, No. 128258.
(42) Shimizu, E.; Zhao, M. R.; Nakanishi, H.; Yamamoto, A.; Yoshida, S.; Takada, M.; Ogura, T.; Sone, S. Differing Effects of staurosporine and UCN-01 on RB Protein Phosphorylation and Expression of Lung Cancer Cell Lines. *Oncology* **1996**, *53* (6), 494−504.

https://doi.org/10.1021/acsomega.3c07338
ACS Omega 2024, 9, 4528−4539