

RESEARCH

Open Access



Fused in silico and bioactivity evaluation method for drug discovery: T001-10027877 was identified as an antiproliferative agent that targets EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R}

Linxiao Wang^{1*}, Xiaoling Huang¹, Shidi Xu¹, Yufeng An¹, Xinya Lv¹, Wufu Zhu^{1*}, Shan Xu¹, Yuanbiao Tu², Shuhui Chen^{3*}, Qiaoli Lv³ and Pengwu Zheng¹

Abstract

Background Facing the significant challenge of overcoming drug resistance in cancer treatment, particularly resistance caused by mutations in epidermal growth factor receptor (EGFR), the aim of our study was to identify potent EGFR inhibitors effective against the ^{T790M/C797S/L858R} mutant, a key player in resistance mechanisms.

Methods Our integrated in silico approach harnessed machine learning, virtual screening, and activity evaluation techniques to screen 5105 compounds from three libraries, aiming to find candidates capable of overcoming the resistance conferred by the T790M and C797S mutations within EGFR. This methodical process narrowed the search down to six promising compounds for further examination.

Results Kinase assays identified three compounds to which the T790M/C797S/L858R mutant exhibited increased sensitivity compared to the T790M/L858R mutant, highlighting the potential efficacy of these compounds against resistance mechanisms. Among them, **T001-10027877** exhibited dual inhibitory effects, with IC₅₀ values of 4.34 μM against EGFR^{T790M/C797S/L858R} and 1.27 μM against EGFR^{T790M/L858R}. Further investigations into the antiproliferative effects in H1975, A549, H460 and Ba/F3-EGFR^{L858R/T790M/C797S} cancer cells revealed that **T001-10027877** was the most potent anticancer agent among the tested compounds. Additionally, the induction of H1975 cell apoptosis and cell cycle arrest by **T001-10027877** were confirmed, elucidating its mechanism of action.

Conclusions This study highlights the efficacy of combining computational techniques with bioactivity assessments in the quest for novel antiproliferative agents targeting complex EGFR mutations. In particular, **T001-10027877**

*Correspondence:

Linxiao Wang
wanglinxiao542@163.com
Wufu Zhu
zhuwufu-1122@163.com
Shuhui Chen
chenshuhui2008@126.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

has great potential for overcoming EGFR-mediated resistance and merits further *in vivo* exploration. Our findings contribute valuable insights into the development of next-generation anticancer therapies, demonstrating the power of an integrated drug discovery approach.

Keywords EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R}, Machine learning, Virtual screening, Bioactivity evaluation, Antiproliferative agent

Introduction

Epidermal growth factor receptor (EGFR) is a transmembrane protein in the tyrosine kinase receptor family [1]. When the extracellular domain of EGFR binds to its ligands (such as EGF or TGF- α), it leads to receptor dimerization and activation, thereby triggering the intracellular tyrosine kinase activity of EGFR [2]. This activates a series of downstream signalling pathways, such as the MAPK, PI3K/Akt, and JAK/STAT pathways, which regulate cell growth, proliferation, migration, and survival [3].

Aberrant activation of the EGFR signalling pathway is associated with the occurrence, progression, and responsiveness of various cancers to treatment [4]. For example, genetic mutations or amplifications of EGFR are common in non-small cell lung cancer and are associated with EGFR sensitivity to tyrosine kinase inhibitors [3]. In addition, other factors related to the EGFR signalling pathway, such as miR-20b [5], HOXA3 [6], and JMJD8 [7], have been found to play roles in cancer that are related to the activation of the EGFR signalling pathway.

To inhibit excessive activation of the EGFR signalling pathway, fourth-generation EGFR inhibitors have been developed (Fig. 1). First-generation EGFR inhibitors (such as erlotinib and gefitinib) primarily target activated EGFR mutants, although the T790M mutant exhibits resistance to these inhibitors [8], and can cause toxic side effects when targeting wild-type EGFR [9]. Second-generation EGFR inhibitors (such as afatinib and dacomitinib) are designed to target the T790M mutation. However, they may lead to more side effects [10, 11] (such as rashes and gastrointestinal disorders) due to their inhibition of wild-type EGFR, which has limited their further clinical application [12]. Third-generation EGFR inhibitors (such as osimertinib) specifically target the T790M mutation and have weaker inhibitory effects on wild-type EGFR, and therefore, fewer side effects [13–15]. Fourth-generation EGFR inhibitors, currently in clinical stages, primarily target triple mutants, including C797S (such as Del19/T790M/C797S and L858R/T790M/C797S) [16]. In conclusion, the discovery of novel fourth-generation EGFR inhibitors has significant research implications for overcoming triple mutants mediated by C797S [17].

In recent years, machine learning techniques have been widely applied in various fields, such as biology, medicine, and computer science. In the field of drug discovery, machine learning has also been extensively used; for

example, to overcome drug resistance using machine learning models [18], the broad-spectrum active compound ruthenium(II) N-heterocyclic carbenes [19], diarylurea antibiotics [20] and anticancer drugs [21] have been discovered.

This study utilized ROC-guided machine learning, virtual screening, and bioactivity evaluations to discover potential compounds targeting EGFR^{T790M/C797S} and EGFR^{T790M/C797S/L858R}. The DUD-E website was used to construct training datasets and evaluate three machine learning models, and the best model was ultimately selected for potential compound prediction. Receptor selection, virtual screening, cluster analysis, and binding analysis were performed, and six promising compounds were identified. After conducting kinase assays, antiproliferative activity tests, AO (Acridine Orange) staining, cell cycle experiments, qRT-PCR, and haemolytic activity evaluations, Compound **T001-10027877** was identified as a novel EGFR^{T790M/C797S}/EGFR^{T790M/C797S/L858R} inhibitor. The entire framework of the study is illustrated in Fig. 2.

Methods and materials

Data preparation

Dataset preparation for machine learning

First, 100 known active compounds (Table S1) were obtained from Selleck (<https://www.selleck.cn/>) and used as a reference to ensure the quality of the datasets used for machine learning. Then, based on the SMILES files of the active compounds, decoys were obtained according to the DUD-E website [22] (<https://dude.docking.org/>). The DUD-E database processed these files and returned a pool of decoy molecules. From this pool, we methodically selected one molecule out of every seven, ultimately forming a set of 220 decoys. The active dataset (100 compounds) and the decoy dataset (220 compounds) were amalgamated to create a dataset, encompassing a total of 320 compounds (Table S2). This consolidated dataset was subsequently converted into an SDF file to facilitate advanced analysis of its physicochemical properties.

Moreover, the physicochemical properties of the active and inactive molecules were predicted using KNIME [23] (<https://www.KNIME.com/>). This process (Figure S1) utilized modules such as SDF Reader, RDKit From Molecule, RDKit Descriptor Calculation, and CSV Writer to predict the physicochemical properties of the 320 compounds. Notably, before the predictions were made, a

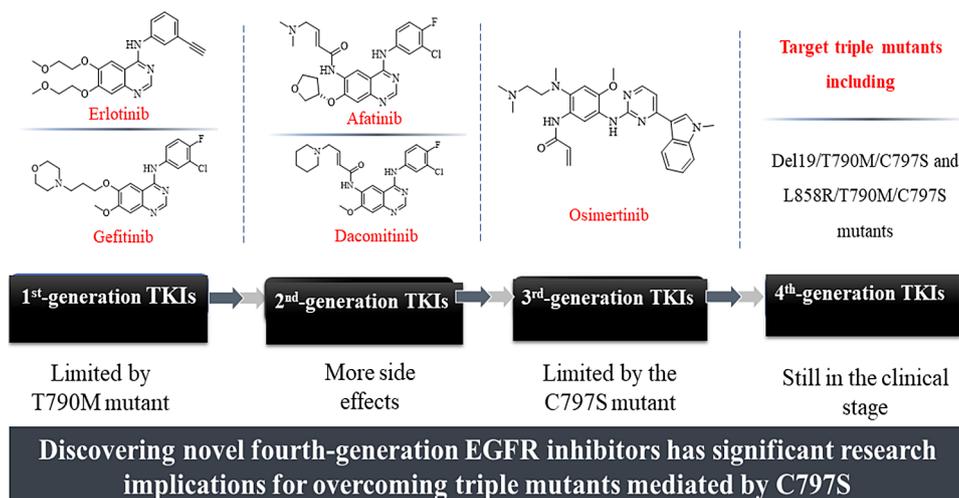


Fig. 1 Features of fourth-generation EGFR inhibitors

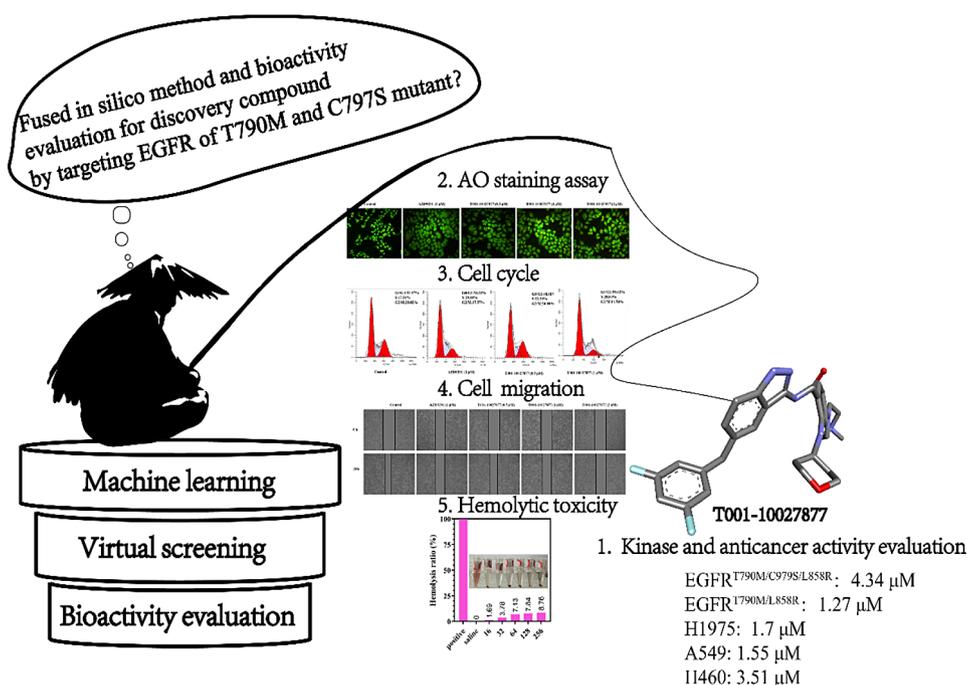


Fig. 2 The machine learning, virtual screening and bioactivity evaluation framework

“class” row was added for classification, where the active molecules were given a label of 1 and inactive molecules were given the label 0. This step is crucial for ensuring the accuracy and reliability of both data processing and the resultant predictions.

Compound library preparation

First, three compound libraries (including the FDA, Bioactivity and Specs libraries; over 220000 compounds) were uploaded into the cheminformatics software KNIME to remove the same molecules and then to predict the physicochemical properties of those remaining (a similar method was described in section “Dataset

preparation for machine learning).” These physicochemical properties were then used as features for further machine learning to predict potential compounds.

Machine learning and the prediction of potential active compounds

Selecting the best-performing machine learning model

Using the data mining and analysis software Orange3 [24], we evaluated five machine learning models (KNN, SVM, random forest, neural network and gradient boosting) on the training datasets in section “Dataset preparation for machine learning”. The machine learning parameters were set as follows: (1) 70% of the datasets

were used for training, and 30% were used for testing; (2) the number of cross-validations was set to 10-fold; and (3) the other parameters were kept at their default values. We considered AUC, CA, FA, precision, and recall to select the best-performing machine learning model for predicting active compounds.

Predicting potential active compounds with the random forest model

For this section, we utilized the top-performing machine learning model, random forest, to forecast the per-compound “class” values for the compounds in the FDA library and Specs library. It is crucial to clarify that these scores, referred to as “class” values, do not carry any units. These class values represent the likelihood of a compound being classified within a specific category, with scores ranging from 0 to 1. Our primary goal was to pinpoint compounds whose “class” values were close to 1 or exceeded 0.5, which indicated a higher probability of relevance to our research objectives. Compounds meeting these criteria were subsequently identified as significant findings for the forthcoming phase of our virtual screening study.

Virtual screening for hit compounds

Docking software and protein selection

Recent research [25] has shown that Vina boasts comparable performance to other open source and academic software programs, including LeDock [26], Gold [27], Moe [28], and Glide [29]. Furthermore, Vina has proven to be a valuable and potent tool in numerous docking research projects aimed at identifying new active compounds. As a result, Vina was chosen as the docking software for this virtual screening.

Our primary aim was to identify compounds capable of targeting EGFR mutants, specifically those with dual or triple mutations. This objective necessitates the careful selection of proteins for virtual screening to ensure the identification of highly specific and effective inhibitors. Therefore, three proteins, 5ZWJ (EGFR^{T790M/C797S/V948R} mutant), 6JRJ (EGFR^{T790M/C797S} mutant) and 6LUD [30–32] (EGFR^{L858R/T790M/C797S} mutant), were evaluated to select suitable proteins for virtual screening. Using these three targets, the original protein ligands were extracted so that they could be docked into the active site again. The RMSD and alignment model were used to evaluate

the sensitivity of Vina towards these three proteins. Finally, the complex with the best RMSD and binding affinity was selected as the docking acceptor for the virtual screen. The docking parameters are listed in Table 1. The affinity, determined from the docking results, and the aligned model were used as rules to select suitable docking receptors.

Compound normalization and preparation

In section “[Compound library preparation](#)”, we detailed the normalization process for hit compounds identified through machine learning using KNIME software. This workflow (Figure S2) includes several modules: File Reader, RDKit Canon SMILES, Duplicate Row Filter, RDKit Add Hydrogens, RDKit Generate Coordinates, RDKit Optimize Geometry, and SDF Writer. These steps systematically prepare the hit compounds by adding hydrogens, generating molecular coordinates, optimizing geometries, and ultimately saving the data in SDF format. This meticulous preparation culminates in converting the compound library into PDBQT files, making them ready for subsequent applications.

Virtual screening

First, the protein was processed in Discovery Studio 3.0 to remove water molecules, ions, and the original ligand and to correct the residues. Subsequently, the PDBQT formats of both the protein and molecules were utilized as input according to the parameters specified in the configuration file outlined in section “[Docking software and protein selection](#)”. Next, the compounds with the highest affinity were preserved for cluster analysis using Discovery Studio 3.0. Ultimately, the potential molecules were chosen based on analysis of both the cluster library and binding model. Detailed methodological details of the virtual screening can be found in our previous work [33–35].

EGFR kinase assay

The kinase inhibition assays were conducted by Heyan Biopharmaceutical Technology Co., Ltd. (Wuhan). The compounds were tested for their inhibitory activity against EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R} using the ADP-GLO and Lance Ultra methods. Detailed methodological details can be found in our previous work [36].

Antitumour evaluation

To initiate the experiment, four cell lines (A549, H1975, H460 and Ba/F3-EGFR^{L858R/T790M/C797S}) in the logarithmic phase of growth were carefully placed into 96-well plates and allowed to incubate for 24 h (the A549, H1975, H460 and Ba/F3-EGFR^{L858R/T790M/C797S} cells used in our study were obtained from Heyan Biopharmaceutical Technology Co., Ltd. (Wuhan), as per the guidelines and

Table 1 Docking parameters of the proteins ZWJ, 6JRJ and 6LUD

Protein ID	Central coordinates (x, y, z)	Grid box Size (x, y, z)	Num modes	Other parameters
5ZWJ	6.77, 21.42, 12.09	30, 30, 30	9	Default
6JRJ	-51.36, 0.69, -22.60	30, 30, 30	9	Default
6LUD	-49.73, -0.039, -18.22	30, 30, 30	9	Default

protocols approved for research). Subsequently, 20 μL of compound-containing medium was evenly distributed into each well, with five different concentrations of each compound and three replicates for each concentration. The final concentrations of the six compounds, D008-10022050, D008-10206173, D008-10099272, D008-10135583, T001-10026427, and **T001-10027877**, were 100.0, 33.3, 11.1, 3.7, and 1.2 μM . After 72 h of incubation, MTT solution (5 mg/ml) was added to each well, and the plates were allowed to incubate for four hours. Once the incubation period ended, the MTT solution was removed, and DMSO (150 μl) was added to each well. Finally, an enzyme marker was used to measure the optical density (OD) of each well at 492 nm.

To determine the antiproliferative effects of the compounds on tumour cells, the measured optical density values were used. The inhibition rate (%) was calculated using the following formula: $(\text{OD control} - \text{OD treatment}) / (\text{OD control} - \text{OD blank}) \times 100\%$. Using SPSS software, we calculated the half-maximal inhibitory concentration (IC_{50}) for each of the compounds on the tumour cells.

AO staining analysis

Acridine orange (AO) is a dye that interacts with DNA and RNA by intercalation or electrostatic attraction [37]. In living cells, AO primarily stains the nucleus green; in dead or damaged cells, AO can penetrate the cell and bind to DNA, staining the nuclei of dead cells orange or red and thus differentiating between living and dead cells. For this assay, T001-10027877 was administered at concentrations of 0.5, 1, and 2 μM , with 1 μM AZD9291 serving as the positive control and untreated cells serving as the blank control. The entire experimental method has been described in our previous research [38].

Cell cycle

H1975 cells were seeded at a density of 4.0×10^5 cells per flask and cultured at 37 $^{\circ}\text{C}$ in a 5% CO_2 incubator for 24 h. The drugs were dissolved in 20 μL of DMSO and then diluted in culture medium to the desired concentrations (AZD9291: 1 μM ; and **T001-10027877**: 0.5 μM or 1 μM) before being added to the culture flasks. After incubation, the cells were then digested with trypsin, centrifuged at 1000 rpm for 5 minutes to remove the supernatant, and washed twice with PBS. Next, 10 mL of 70% ethanol was added at 4 $^{\circ}\text{C}$ to fix the cells for more than 24 h. The fixed cells were centrifuged at 2000 rpm for 5 minutes, the supernatant was discarded, and the cells were washed twice with PBS before 300 μL of PI staining solution was added. After incubation in the dark at 4 $^{\circ}\text{C}$ for 30 min, the cell cycle distribution was analysed using flow cytometry, and key data were retained.

Migration experiment

To conduct the cell scratch assay, log-phase H1975 cells were seeded into a six-well plate, with each well receiving approximately 5×10^5 cells. Before seeding, parallel and equidistant lines were marked on the underside of the plate to serve as guides for the scratch. After 24 h, the cells had reached confluence, and a pipette tip was used to create a precise scratch along the premarked lines. Subsequently, the plate was rinsed twice with PBS to remove any detached cells, and fresh culture medium was added. At this stage, the cells were treated with **T001-10027877** at concentrations of 0.5, 1, and 2 μM , and a control group was treated with 1 μM AZD9291. Initial observations and photographs of the scratched area were taken under a microscope to establish baseline data (0 h). Culture was then continued in the incubator for 24 h. After the plate was retrieved, the cells were observed and photographed under a microscope again to document the migration and healing process, with particular attention given to the effects of **T001-10027877** and AZD9291 on cell migration.

Real-time PCR

Using 1 μM AZD9291 as a positive control, fluorescence quantitative PCR was utilized to investigate the concentration-dependent inhibition of the substances on EGFR and mTOR in H1975 cells. The concentrations of T001-10027877 used were 0.5 and 1 μM . The procedure was described by Sun et al. [38].

Haemolytic activity assay

A haemolytic activity assay was conducted to evaluate the potential haemolytic effects of **T001-10027877**. The concentrations tested for **T001-10027877** included 0 (as a control), 16, 32, 64, 128, and 256 $\mu\text{g/mL}$. In this assay, 1% Triton X-100, which is known to cause 100% haemolysis of sheep blood erythrocytes, was used as a positive control to indicate complete cell lysis. Saline solution served as a negative control, representing no haemolytic activity. The specific experimental method was based on our previous research [38].

Results

Machine learning to predict potential compounds

Selection of the machine learning model

To select the best machine learning model, the ROC method was used. First, five machine learning models (KNN, SVM, random forest, neural network and gradient boosting) were built with Orange3 to evaluate the training set to identify the best model, and the results are described in Table 2; Fig. 3. The AUC values can be calculated from Table S3. According to Table 2; Fig. 3, the AUCs of the random forest (0.98) and SVM (0.98) models were better than those of the other three models (KNN

Table 2 Performance of the five machine learning models

Model	AUC	CA	F1	Precision	Recall
KNN	0.93	0.88	0.88	0.89	0.88
SVM	0.98	0.95	0.95	0.95	0.95
Random Forest	0.98	0.97	0.97	0.97	0.97
Neural Network	0.97	0.95	0.95	0.95	0.95
Gradient Boosting	0.97	0.93	0.93	0.93	0.93

(0.93), neural network (0.97) and gradient boosting (0.97)). Among the random forest and SVM models, the random forest model also performed better in terms of CA, F1, precision and recall. Based on these results, the random forest model was selected as the machine learning model in this study.

Using the random forest model to predict potential compounds

After the physicochemical properties of the compounds in the library were predicted, the data mining software Orange3 was used to construct a random forest model to predict the probability of target inhibition. Compounds with a positive rate greater than 0.5 were retained, which resulted in approximately 5105 compounds (Table S4). Then, these compounds were processed by KNIME to add hydrogen atoms, generate coordinates and optimize the geometry and saved as an SDF file. Finally, the SDF file was converted into PDBQT by Open Babel 3.0 for further study. The workflow is shown in Fig. 4.

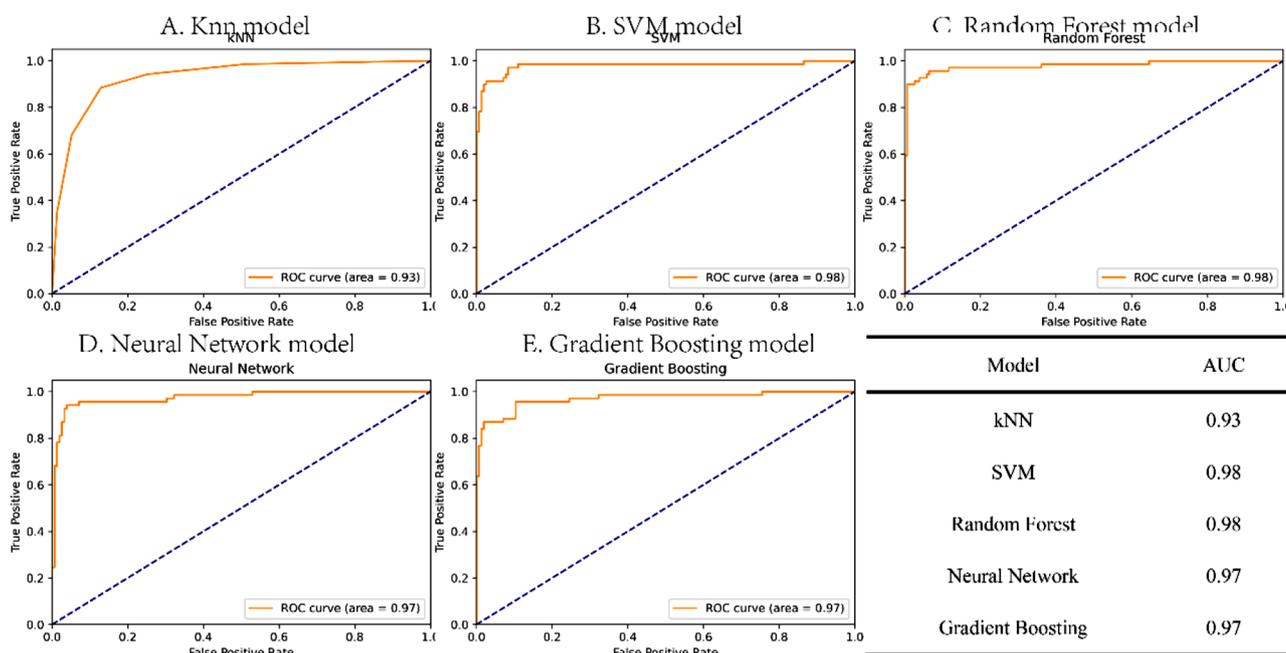


Fig. 3 ROC curves of five machine learning models: KNN (A), SVM (B), random forest (C), neural network (D) and gradient boosting (E). The AUCs of the five machine learning methods are summarized in Fig. 3F

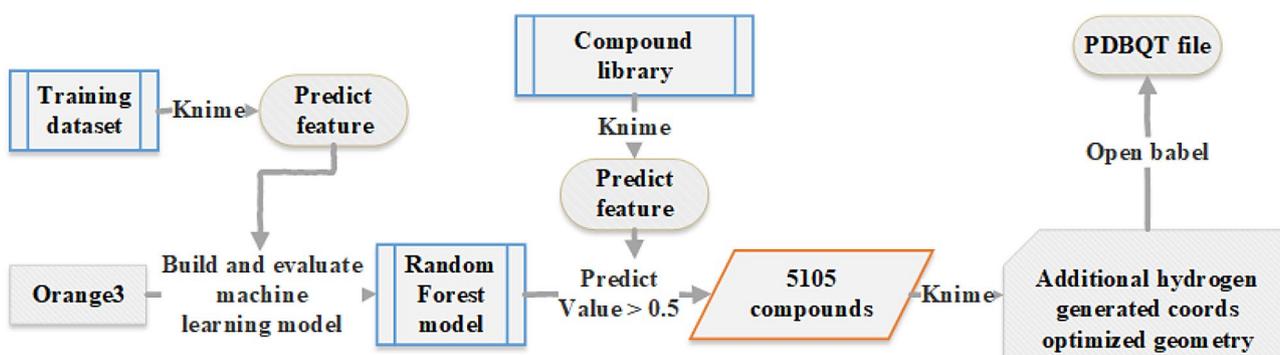


Fig. 4 Workflow of the random forest model prediction and form conversion of the compound library

Table 3 RMSDs after redocking the original ligand relative to that of the original protein

Protein	Affinity (kcal/mol)	RMSD of the aligned original ligand
5ZWJ	-8.7	0.35
6JRJ	-11.4	0.18
6LUD	-8.0	0.07

Virtual screening

Selecting suitable proteins for virtual screening

To select suitable virtual screening software, the three mutant proteins were selected as receptors, and the original ligand of each was redocked into the active pocket to evaluate the RMSD. Thus, the three proteins were used as receptors, and AutoDock Vina was used to dock the original ligand into the active pocket to evaluate the sensitivity of Vina towards the three proteins. As shown in Table 3; Fig. 5, the original ligand in 6JRJ showed better affinity (affinity value: -11.4 kcal/mol) than those in 5ZWJ (affinity value: -8.7 kcal/mol) and 6LUD (affinity value: -8.0 kcal/mol). In addition, among the three ligands, the RMSD of the aligned original ligand of 6JRJ (RMSD: 0.18) was better than that of 5ZWJ (RMSD: 0.35) but close to that of 6LUD (RMSD: 0.07). Taking these results into consideration, clearly, protein 6JRJ was more suitable for virtual screening than 5ZWJ and 6LUD.

3.2.2 Virtual screening and hit compound selection

After selecting suitable docking software and proteins, approximately 5105 compounds were docked into the active site of the protein 6JRJ, and after setting a threshold affinity of less than -9.5 kcal/mol, 284 compounds remained (Table S5). Then, cluster analysis was performed to obtain 45 compounds based on the diversity of the skeletons. Finally, based on the binding model compared with the original ligand of 6JRJ, 6 compounds were selected as hit compounds, as displayed in Table 4. In addition, the 6 selected compounds were aligned with the original ligand of 6JRJ, and nearly all showed conformations more similar to the original ligand than did

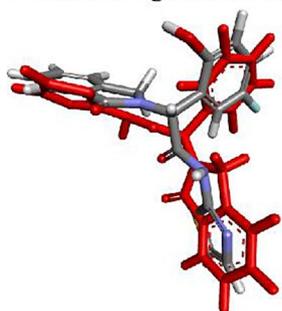
the 39 unselected compounds, as displayed in Fig. 6. These results indicated that the selected compounds may inhibit EGFR^{T790M/L858R} and EGFR^{T790M/C797S/L858R}. Finally, all selected compounds were purchased from TargetMol (US).

EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R} inhibition assays and binding model analysis

To prove our suspicions, kinase assays were performed to evaluate whether the hit compounds were on target. Table 5 shows that EGFR^{T790M/C797S/L858R} exhibited increased sensitivity to nearly all the compounds. For example, three compounds (D008-10099272, D008-10135583 and T001-10027877) inhibited EGFR^{T790M/C797S/L858R} with IC₅₀ values of 3.26, 3.25 and 4.32 μM, respectively, and only one compound inhibited EGFR^{T790M/L858R} (T001-10027877: 1.27 μM). Additionally, compared to AZD9291, which served as a positive control with IC₅₀ values of 0.64 and 0.012 μM against EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R}, respectively, T001-10027877 clearly demonstrated less potent kinase inhibition. However, the unique structural architecture of T001-10027877 offers significant potential for future structural modifications. Therefore, T001-10027877 emerged as a promising lead compound for devising strategies to counteract EGFR-induced resistance, especially by targeting the T790M and C797S mutations.

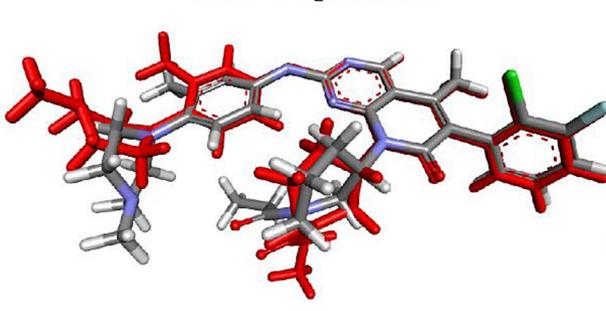
In addition, T001-10027877 was selected for binding analysis, using D008-10206173 as the negative control, as shown in Fig. 7. According to recent research, the residues M790, L858, S797 and M793 are important for EGFR inhibitors to overcome resistance mediated by T790M or C797S. Figure 7A shows the robust inhibitory activity of AZD9291 against EGFR resistance, as evidenced by the formation of four hydrogen bonds with the crucial amino acids K745, S797, and M793 and the establishment of five pi-cation interactions with other amino acids. This demonstrates its high efficacy in targeting the EGFR protein. On the other hand, Fig. 7B shows

A Redock of ligand of 5ZWJ



RMSD: 0.35

B Redock of ligand of 6JRJ



RMSD: 0.18

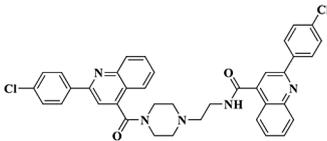
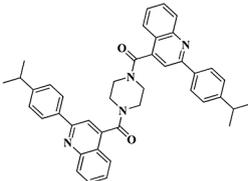
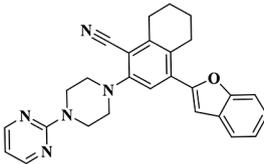
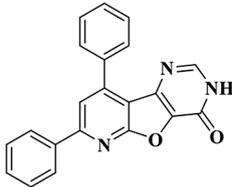
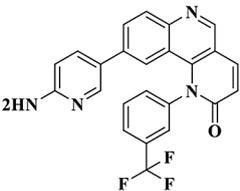
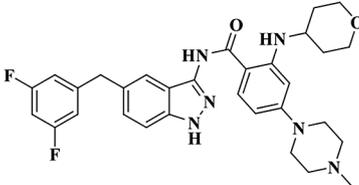
C Redock of ligand of 6LUD



RMSD: 0.07

Fig. 5 The original ligands (red) aligned with the redocked ligands

Table 4 Docking information for the 6 selected compounds

Name	Structure	Affinity (kcal/mol)	Cluster
D008-10022050		-11	17
D008-10206173		-11.1	17
D008-10099272		-11.3	17
D008-10135583		-10.8	10
T001-10026427		-11	36
T001-10027877		-9.9	42

that **T001-10027877** exhibits a diverse set of interactions, including hydrogen bonding, hydrophobic, and pi-cation interactions, notably enhancing its binding to the S797 residue via additional interactions with M790, S796, and M793. Despite this broad interaction spectrum, the potency of **T001-10027877** did not reach that of AZD9291.

Further analysis of D008-10206173 (presented in Fig. 7C) shows that this compound forms primarily hydrophobic and pi-stacking interactions, which are

inherently weaker than hydrogen bonds. This comparison revealed that **T001-10027877** demonstrates greater activity than D008-10206173 but remains less potent than AZD9291. The similar interaction patterns of **T001-10027877** and AZD9291 with the protein, particularly in overcoming EGFR resistance caused by the T790M or C797S mutation, suggest that **T001-10027877**, despite its reduced potency compared to that of AZD9291, still holds potential as a candidate for overcoming resistance mediated by specific EGFR mutants. Such integrated

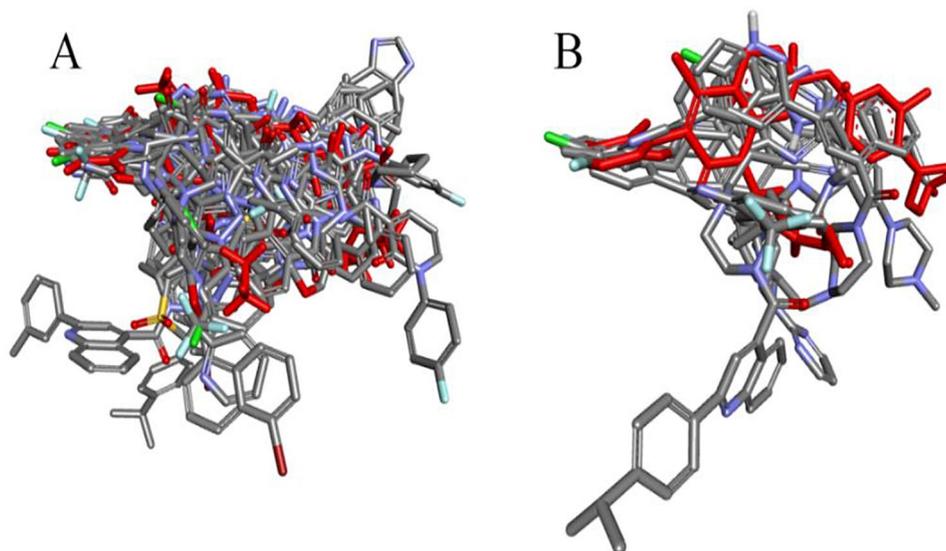


Fig. 6 (A) The original ligand of protein 6JRJ aligned with the 39 unselected compounds. (B) The original ligand of protein 6JRJ aligned with the 6 selected compounds. It is important to highlight that these compounds were chosen from among 45 candidates identified through a clustering process

Table 5 Kinase inhibitory activity of six selected compounds against EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R}

Compd.	IC ₅₀ (μM)	
	EGFR ^{T790M/C797S/L858R}	EGFR ^{T790M/L858R}
D008-10022050	>10.0	>10.0
D008-10206173	>10.0	>10.0
D008-10099272	3.26	>10.0
D008-10135583	3.35	>10.0
T001-10026427	>10.0	>10.0
T001-10027877	4.34	1.27
AZD9291 ^a	0.64	0.012

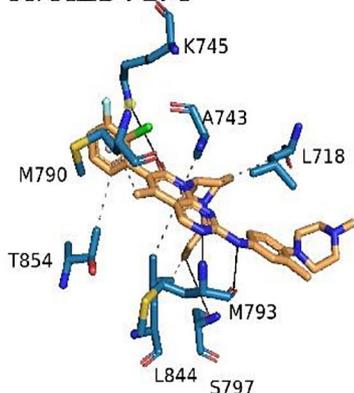
^a Used as the positive control

analysis emphasizes the nuanced efficacy of these compounds in targeting EGFR mutations, highlighting the ability of T001-10027877 to serve as an alternative for combating EGFR resistance.

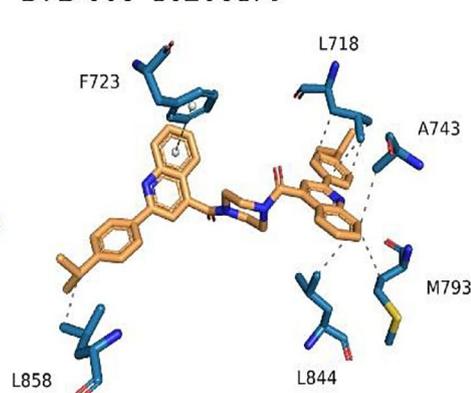
Antiproliferative activity

Furthermore, the antiproliferative activity of the six selected compounds was evaluated by the MTT method with four cancer cell lines (A549, H1975, H460 and Ba/F3-EGFR^{L858R/T790M/C797S}). These cell lines were chosen to encompass a broad spectrum of EGFR mutation statuses: A549 and H460 as wild-type EGFR controls, H1975 representing the EGFR^{T790M/L858R} mutations associated with acquired resistance to first-generation EGFR tyrosine kinase inhibitors, and Ba/F3-EGFR^{L858R/T790M/C797S} to model compound efficacy against the complex scenario involving the T790M, L858R, and C797S mutations. As shown in Table 6, all the compounds inhibited the growth of A549, H1975, H460 and Ba/F3-EGFR^{L858R/T790M/C797S} cells to different degrees, with IC₅₀ values ranging from 1 to 50 μM. Moreover, nearly all the compounds were

A. AZD9291



B. D008-10206173



C. T001-10027877

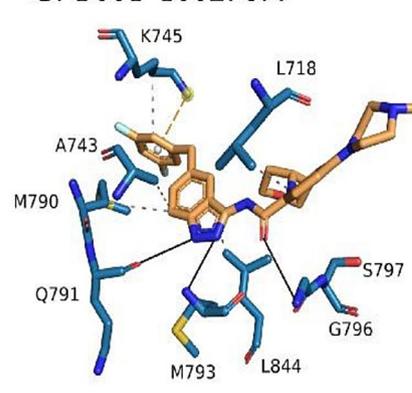


Fig. 7 Analysis of the molecular interactions of AZD9291-6JRJ, T001-10027877-6JRJ and D008-10206173-6JRJ

Table 6 The antiproliferative activities of the six selected compounds in A549, H460 and Ba/F3-EGFR^{L858/T790M/C797S} cells

Compd.	IC ₅₀ (μM ± SD)			
	H1975	A549	H460	Ba/F3-EGFR ^{L858/T790M/C797S}
D008-10022050	38.17 ± 3.56	> 50	> 50	> 50
D008-10206173	> 50	> 50	> 50	> 50
D008-10099272	41.97 ± 4.69	> 50	> 50	> 50
D008-10135583	39.92 ± 4.05	15.90 ± 2.73	4.68 ± 0.41	> 50
T001-10026427	43.14 ± 4.68	34.68 ± 3.61	44.22 ± 4.68	> 50
T001-10027877	1.70 ± 0.12	1.55 ± 0.13	3.51 ± 0.36	4.46 ± 0.46
AZD9291 ^a	4.951 ± 0.51	4.47 ± 0.45	5.38 ± 0.49	4.34 ± 0.34

^a Used as the positive control

more sensitive to the H1975 cell line, which implies that these selected compounds could be used as potential compounds for overcoming resistance. Notably, the **T001-10027877** showed 1- to 3-fold greater activity than the positive control compound AZD9291 (4.95, 1.55, 5.38 and 4.34 μM) against A549, H1975, H460 and Ba/F3-EGFR^{L858/T790M/C797S} cells, with IC₅₀ values of 1.7, 1.55, 3.51 and 4.46 μM, respectively. Notably, only **T001-10027877** and AZD9291 inhibited the Ba/F3-EGFR^{L858/T790M/C797S} cell line, with IC₅₀ values of 4.46 μM and 4.34 μM, respectively. These data indicate that compared to AZD9291, **T001-10027877** shows similar or even superior inhibition of Ba/F3-EGFR^{L858/T790M/C797S} cells harbouring the triple mutation, as well as the double-mutant H1975 cells. This finding highlights the potential of **T001-10027877** inhibiting the proliferation of cancer cells with the T790M and C797S mutations in EGFR.

AO staining assays

Our goal was to identify a potential compound to overcome the resistance induced by the EGFR mutations T790M and C797S. Moreover, **T001-10027877** demonstrated greater inhibition of the EGFR T790M mutant

than the C797S mutant. Consequently, the H1975 cell line was chosen for assessing the antiproliferative effects using the AO staining assay at AO concentrations of 0.5, 1, and 2 μM (Fig. 8). As shown in Fig. 8, the H1975 cells in the control group exhibited a normal morphology, emitting circular and dense green fluorescence. Notably, both the positive control (AZD9291) and **T001-10027877** induced significant shrinkage of the cell membrane and the appearance of sharply defined cell edges, indicating the induction of apoptosis. At a concentration of 2 μM, **T001-10027877** induced apoptosis to the same extent as AZD9291. Moreover, as the concentration of **T001-10027877** increased, its ability to induce H1975 cell apoptosis also increased.

Effects of T001-10027877 on H1975 cell cycle progression

To explore the impact of Compound **T001-10027877** on cell proliferation, cell cycle analysis was conducted in H1975 cells. As shown in Fig. 9, Compound **T001-10027877** induced cell cycle arrest at the G0/G1 checkpoint in a dose-dependent manner compared to that in the control group. As the concentration of Compound **T001-10027877** increased, the S phase population increased from 17.01 to 28.84%, while the G2/M phase population decreased from 29.02 to 11.54%. The proportion of cells in G0/G1 phase did not significantly change. Additionally, at a concentration of 1 μM, Compound **T001-10027877** caused cell cycle arrest at the S checkpoint at a rate of 28.84%, which is similar to that of the positive control AZD9291 (29.10%).

Inhibitory effects of T001-10027877 on H1975 cell migration

A cell scratch assay was used to investigate the effect of **T001-10027877** on the migratory capacity of H1975 cells. The cells were treated with 0.5, 1 or 2 μM **T001-10027877** for 24 h and then photographed under a microscope. As shown in Fig. 10, after 24 h of culture, the cells in the control group significantly migrated towards the centre. The cells treated with **T001-10027877** migrated less than those in the control group, and this inhibitory effect was more potent at 2 μM **T001-10027877** than at 0.5 μM **T001-10027877**. The

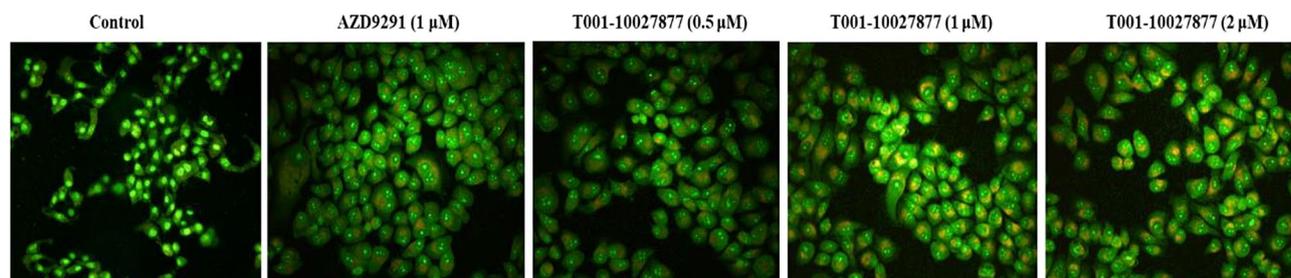


Fig. 8 Cell morphology observed by fluorescence microscopy after treatment with **T001-10027877**

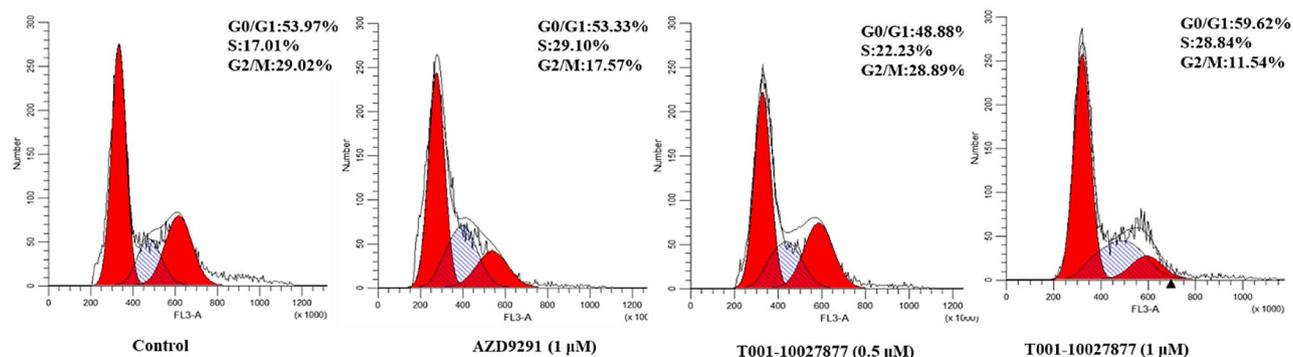


Fig. 9 Compound **T001-10027877** induces H1975 cell cycle arrest in a dose-dependent manner

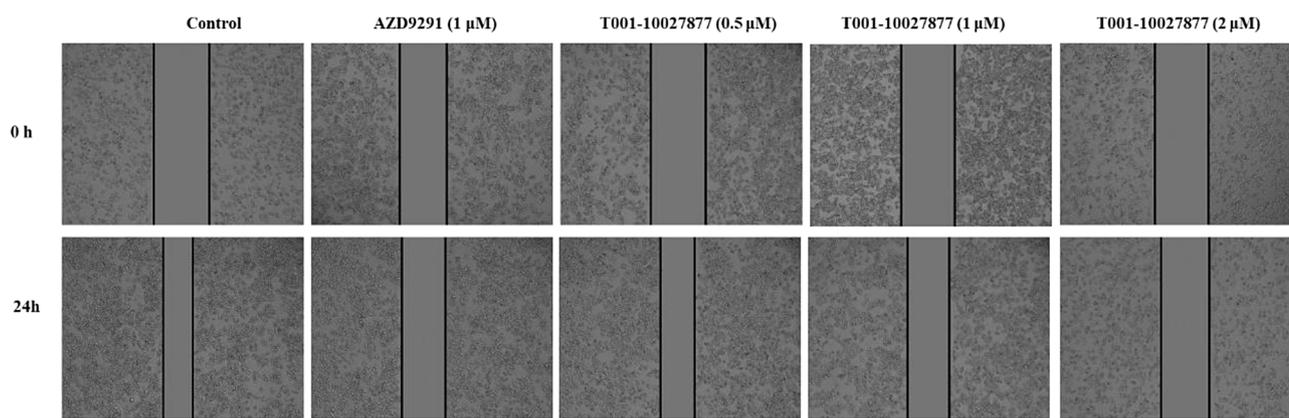


Fig. 10 Dose-dependent inhibition of H1975 cell migration by Compound **T001-10027877**: A comparative study with AZD9291

inhibitory effect of **T001-10027877** was similar to that of AZD9291 at a concentration of 1 μ M. These results demonstrated that **T001-10027877** had a dose-dependent inhibitory effect on the migratory capacity of H1975 cells and exhibited efficacy similar to that of the positive control AZD9291.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis after treatment with **T001-10027877**

To determine the mechanism of action of Compound **T001-10027877** in tumour cells, fluorescence quantitative PCR was used to examine the EGFR and mTOR expression levels in H1975 cells. In untreated H1975 cells, the expression levels of EGFR and mTOR were at baseline levels, as shown in Fig. 11. Treatment with 0.5 μ M **T001-10027877** effectively reduced the relative EGFR expression to 0.51 and the relative mTOR expression to 0.64. Increasing the concentration of **T001-10027877** to 1 μ M further decreased the relative EGFR expression and mTOR expression to 0.35 and 0.53, respectively, demonstrating dose-dependent inhibition. At 1 μ M, the positive control AZD9291 reduced the relative EGFR expression to 0.28 and that of mTOR to 0.52, slightly surpassing the effects of **T001-10027877** in terms of EGFR suppression

but comparably affecting mTOR expression. These results highlight the potent, dose-responsive inhibitory effect of **T001-10027877** on the EGFR and mTOR signalling pathways, as **T001-10027877** shows marked selectivity for EGFR at relatively high concentrations. Furthermore, **T001-10027877** demonstrated similar selectivity and efficacy towards EGFR as the positive control AZD9291, indicating comparable levels of expression and activity.

Haemolytic activity of **T001-10027877** at various concentrations

Moreover, haemolytic activity was also considered. As shown in Fig. 12, the haemolytic activity values were 0, 1.69, 3.78, 7.13, 7.84 and 8.76 for **T001-10027877** at 0, 16, 32, 64, 128 and 256 μ g/mL, respectively. All of the haemolytic activity values at various concentrations were less than 9%, which means that compared with the control, 1% Triton X, **T001-10027877** was safe in terms of haemolytic activity for the treatment of cancer cells.

Discussion

Herein, we emphasize the multifaceted evaluation of **T001-10027877**, which spans from machine learning model selection to detailed molecular interaction analysis, highlighting the potential of this compound as a novel

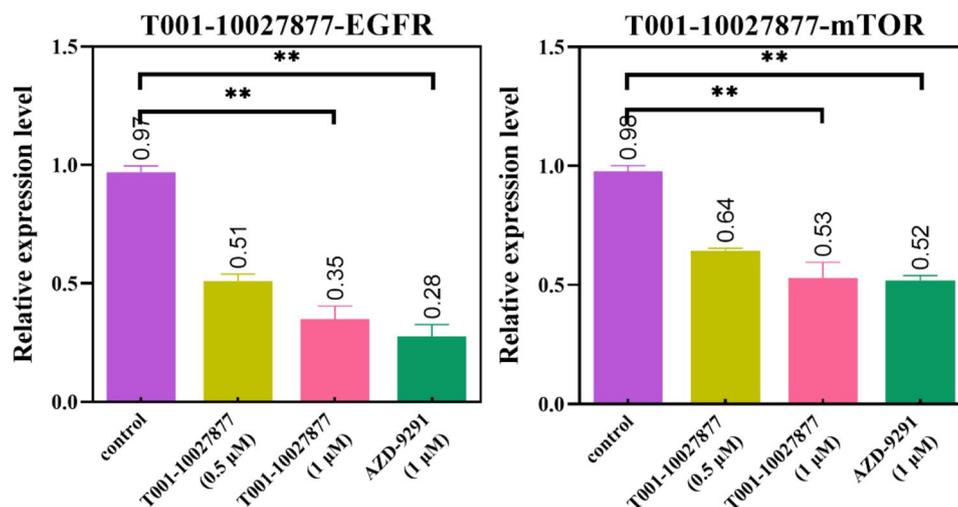


Fig. 11 Quantitative RT-PCR analysis of EGFR and mTOR expression in H1975 cells treated with T001-10027877

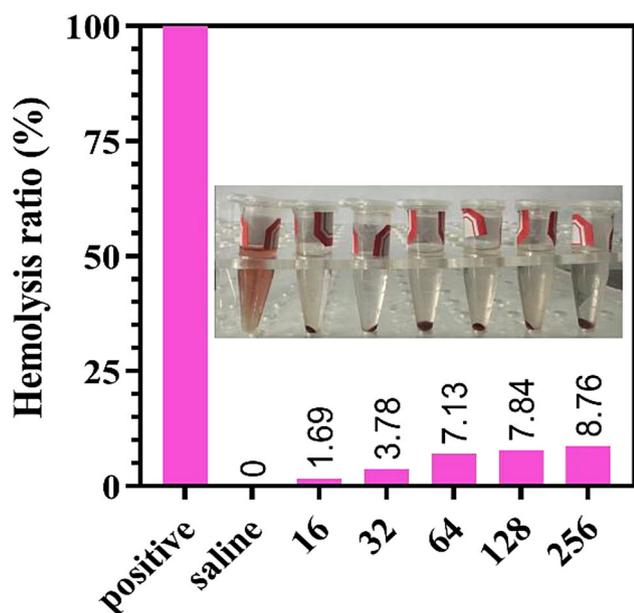


Fig. 12 Evaluation of the haemolytic safety of Compound T001-10027877: A dose-response study

agent for overcoming EGFR-mediated resistance in cancer therapy.

Initially, the selection of the random forest model via the ROC method, as documented in Table 2; Fig. 3, provided a solid foundation for computational screening. This model outperformed the others, including SVM, KNN, neural network, and gradient boosting, by demonstrating superior accuracy in predicting compounds with potential inhibitory activity against EGFR mutants. Such precision led to the identification of approximately 5105 compounds, a testament to the model's efficacy in navigating the complex chemical space of potential inhibitors.

Redocking studies further refined our search, pinpointing protein 6JRJ as the most suitable receptor for virtual

screening due to its affinity for and sensitivity towards the ligands, as shown in Table 3; Fig. 5. This specificity in target selection was crucial for identifying 284 compounds with significant affinity, which were further narrowed down to six hit compounds through cluster analysis, emphasizing the importance of structural diversity in the selection process.

A comparative analysis of these compounds revealed that T001-10027877 had notable inhibitory activity, particularly against the EGFR^{T790M/L858R} and EGFR^{T790M/C797S/L858R} mutants, showing efficacy similar to or superior to that of the positive control, AZD9291. This is supported by molecular interaction studies (Fig. 7A and B), where T001-10027877 exhibited a comprehensive interaction profile with key EGFR residues, suggesting its ability to effectively inhibit EGFR-driven cell proliferation.

Further experimental validation via kinase assays and antiproliferative activity tests across multiple cancer cell lines (A549, H1975, H460, and Ba/F3-EGFR^{L858R/T790M/C797S}) confirmed the potent activity of T001-10027877. Notably, T001-10027877 demonstrated 1- to 3-fold greater activity than did AZD9291 in certain assays, highlighting its potential to serve as a more effective therapeutic agent.

The antimigratory effects of T001-10027877, as evidenced by the cell scratch assay, and its ability to induce cell cycle arrest, as demonstrated via cell cycle analysis, further substantiated the efficacy of this novel compound. These results, together with the fluorescence quantitative PCR findings, revealed that T001-10027877 dose-dependently inhibited EGFR and mTOR expression, revealing its targeted mechanisms of action against cancer cell proliferation.

Finally, the evaluation of haemolytic activity confirmed the safety of T001-10027877, indicating that it is a viable

candidate for further development. The minimal haemolytic effects of this compound, even at high concentrations, align with the requirements for a therapeutically effective yet safe anticancer agent.

In summary, the comprehensive assessment of T001-10027877, from computational predictions to experimental validation, showcases its promise as a novel therapeutic option for overcoming EGFR-mediated resistance. Its potent efficacy, coupled with its favourable safety profile, makes it a compelling candidate for further investigation and potential clinical application in cancer therapy.

Conclusion

In this study, an *in silico* fusion method, which included machine learning, virtual screening and clustering, combined with activity evaluations was used to identify active EGFR inhibitors targeting EGFR^{T790M/C797S/L858R}. First, ROC-guided machine learning was used to identify nearly 5105 compounds from three compound libraries (over 200000 compounds). Second, virtual screening and binding model analysis were performed to obtain six potential compounds. In addition, the kinase assay showed that EGFR^{T790M/C797S/L858R} exhibited greater sensitivity to these six compounds than EGFR^{T790M/L858R}. Among them, Compound T001-10027877 inhibited both EGFR^{T790M/C797S/L858R} and EGFR^{T790M/L858R}, with IC₅₀ values of 4.34 and 1.27 μM, respectively, which means that T001-10027877 could overcome the EGFR resistance mediated by the T790M or C797S mutations. Moreover, the antiproliferative effects of the six hit compounds were also investigated, and T001-10027877 showed the best anticancer activity in the H1975, A549, H460 and Ba/F3-EGFR^{L858R/T790M/C797S} cell lines, with IC₅₀ values of 1.7, 1.5, 3.5 and 4.46 μM, respectively. Furthermore, AO staining and cell cycle experiments revealed that Compound T001-10027877 could induce H1975 cell apoptosis in a concentration-dependent manner. Moreover, the haemolytic activity evaluation suggested that the Compound T001-10027877 is safe for further *in vivo* study.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13065-024-01279-z>.

Supplementary Material 1

Author contributions

LW: Conceptualization, Investigation, Methodology, Formal analysis, Project administration, Data curation, Software, Validation, Writing – original draft, Writing – review & editing, Resources, Funding acquisition. XH: Investigation, Methodology, Visualization, Resources, Data curation, Software, Validation. SX: Visualization, Data curation, Validation. YA: Investigation, Visualization, Validation. XI: Investigation, Visualization, Validation. WZ: Investigation, Methodology, Formal analysis, Project administration, Supervision,

Resources, Funding acquisition. SX: Visualization, Data curation, Validation. YT: Investigation, Methodology, Formal analysis, Project administration, Data curation, Validation, Resources. SC: Conceptualization, Investigation, Methodology, Formal analysis, Project administration, Supervision, Writing – review & editing, Resources, Funding acquisition. QL: Methodology, Formal analysis, Funding acquisition. PZ: Methodology, Formal analysis.

Funding

This work was financially supported by the Jiangxi Province Natural Science Foundation for Youths (20224BAB216112), the Science and Technology Research Foundation of Jiangxi Education Department (GJJ21113), the Ph.D. Startup Foundation of Jiangxi Science and Technology Normal University (2021BSQD25), the “Five-level Progressive” talent cultivation project of Jiangxi Cancer Hospital & Institute (WCDJ2024YQ03) and the Key R&D plan of Jiangxi Province (2021BBG71006).

Data availability

All data generated or analyzed during this study are included in the manuscript. The compound libraries used in this study are accessible via the following links: FDA Library (<https://file.selleckchem.com/downloads/library/20240801-L1300-FDA-approved-Drug-Library.SDF>), Bioactivity Library (<https://file.selleckchem.com/downloads/library/20240801-L1700-Bioactive-Compound-Library-I.SDF> and [https://file.selleckchem.com/downloads/library/20210610-L1700-II-Bioactive-Compound-Library-II-\(Provided-by-Pfizer\)-96well-com.sdf](https://file.selleckchem.com/downloads/library/20210610-L1700-II-Bioactive-Compound-Library-II-(Provided-by-Pfizer)-96well-com.sdf)), and Specs Library (<https://www.specs.net/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Jiangxi Provincial Key Laboratory of Drug Design and Evaluation, School of Pharmacy, Jiangxi Science & Technology Normal University, Nanchang 330013, China

²Cancer Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China

³Jiangxi Provincial Cancer Hospital, Nanchang, Jiangxi 330029, P. R. China

Received: 1 October 2023 / Accepted: 22 August 2024

Published online: 27 August 2024

References

1. Idowu MO. Epidermal growth factor receptor in Lung Cancer the amazing interplay of Molecular Testing and Cytology. *Cancer Cytopathol.* 2013;121(10):540–3.
2. Shao Q, Zhu WL. Ligand binding effects on the activation of the EGFR extracellular domain. *Phys Chem Chem Phys.* 2019;21(15):8141–51.
3. Gandhi J, Zhang JL, Xie Y et al. Alterations in genes of the EGFR Signaling Pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in Lung Cancer Cell Lines. *PLoS ONE.* 2009, 4(2).
4. Tan J, Li H, Ji CL et al. Electron transfer-triggered imaging of EGFR signaling activity. *Nat Commun.* 2022, 13(1).
5. Fu Q, Cheng J, Zhang JD, et al. miR-20b reduces 5-FU resistance by suppressing the ADAM9/EGFR signaling pathway in colon cancer. *Oncol Rep.* 2017;37(1):123–30.
6. Zhang XX, Liu GW, Ding L, et al. HOXA3 promotes tumor growth of human colon cancer through activating EGFR/Ras/Raf/MEK/ERK signaling pathway. *J Cell Biochem.* 2018;119(3):2864–74.
7. Zhang B, Zhang Y, Jiang XZ, et al. JMJD8 promotes malignant progression of Lung Cancer by maintaining EGFR Stability and EGFR/PI3K/AKT pathway Activation. *J Cancer.* 2021;12(4):976–87.

8. Shah R, Lester JF. Tyrosine kinase inhibitors for the treatment of EGFR mutation-positive non-small-cell lung Cancer: a clash of the Generations. *Clin Lung Cancer*. 2020;21(3):E216–28.
9. Karachaliou N, Fernandez-Bruno M, Bracht JWP, et al. EGFR first- and second-generation TKIs—there is still place for them in EGFR-mutant NSCLC patients. *Translational Cancer Res*. 2019;8:S23–47.
10. Fassunke J, Muller F, Keul M et al. Overcoming EGFR(G724S)-mediated osimertinib resistance through unique binding characteristics of second-generation EGFR inhibitors. *Nat Commun*, 2018, 9.
11. Li D, Ambrogio L, Shimamura T, et al. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene*. 2008;27(34):4702–11.
12. Zhang L, Yang YY, Zhou HJ, et al. Structure-activity study of quinazoline derivatives leading to the discovery of potent EGFR-T790M inhibitors. *Eur J Med Chem*. 2015;102:445–63.
13. Dhillon S, Lazertinib. *First Approval Drugs*. 2021;81(9):1107–13.
14. Shi KY, Wang G, Pei JP et al. Emerging strategies to overcome resistance to third-generation EGFR inhibitors. *J Hematol Oncol*, 2022, 15(1).
15. Soria JC, Ohe Y, Vansteenkiste J, et al. Osimertinib in untreated EGFR-Mutated Advanced Non-small-cell Lung Cancer. *N Engl J Med*. 2018;378(2):113–25.
16. Jia Y, Yun CH, Park E, et al. Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature*. 2016;534(7605):129–.
17. Thress KS, Paweletz CP, Felip E, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med*. 2015;21(6):560–2.
18. Choi G, Kim D, Oh J. AI-Based Drug Discovery of TKIs Targeting L858R/T790M/C797S-Mutant EGFR in non-small cell lung Cancer. *Front Pharmacol*, 2021, 12.
19. Al Nasr IS, Koko WS, Khan TA et al. Evaluation of Ruthenium(II) N-Heterocyclic Carbene complexes as enzymatic inhibitory agents with antioxidant, Antimicrobial, Antiparasitical and Antiproliferative Activity. *Molecules*, 2023, 28(3).
20. Catalano A, Iacopetta D, Pellegrino M et al. Diarylureas: repositioning from Antitumor to antimicrobials or Multi-target agents against New Pandemics. *Antibiotics-Basel*, 2021, 10(1).
21. Iwaloye O, Ottu PO, Olawale F et al. Computer-aided drug design in anti-cancer drug discovery: what have we learnt and what is the way forward?. *Inf Med Unlocked*, 2023: 101332–101332.
22. Mysinger MM, Carchia M, Irwin JJ, et al. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for Better Benchmarking. *J Med Chem*. 2012;55(14):6582–94.
23. Berthold MR, Cebon N, Dill F et al. KNIME: The Konstanz Information Miner[C]// KNIME: The Konstanz Information Miner. 31st Annual Conference of the German-Classification-Society, Mar 07–09, GERMAN CLASSIFICAT S, ALBERT LUDWIGS U, Albert Ludwigs Univ, Freiburg, GERMANY. 2008: 319–326.
24. Demsar J, Curk T, Erjavec A, et al. Orange: Data Mining Toolbox in Python. *J Mach Learn Res*. 2013;14:2349–53.
25. Wang Z, Sun H, Yao X, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys*. 2016;18(18):12964–75.
26. Liu N, Xu ZB, Iop. Using LeDock as a docking tool for computational drug design[C]// Using LeDock as a docking tool for computational drug design. International Conference on Civil, Architecture and Disaster Prevention, Oct 19–Nov 04, GUANGDONG ACAD INT ACAD EXCHANGE A U A, ANHUI JIAN-ZHU UNIV S C E, Anhui Jianzhu Univ, Sch Civil Engn, Hefei, PEOPLES R CHINA. 2019,218.
27. Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein-ligand docking using GOLD. *Proteins-Structure Function Genet*. 2003;52(4):609–23.
28. Wang Z, Sun HY, Yao XJ, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys*. 2016;18(18):12964–75.
29. Repasky MP, Murphy RB, Banks JL, et al. Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *J Comput Aided Mol Des*. 2012;26(6):787–99.
30. Kashima K, Kawauchi H, Tanimura H, et al. CH7233163 overcomes osimertinib-resistant EGFR-Del19/T790M/C797S Mutation. *Mol Cancer Ther*. 2020;19(11):2288–97.
31. Shen JY, Zhang T, Zhu SJ, et al. Structure-based design of 5-Methylpyrimidopyridone derivatives as New Wild-Type sparing inhibitors of the Epidermal Growth Factor Receptor Triple Mutant (EGFR^{L858R/T790M/C797S}). *J Med Chem*. 2019;62(15):7302–8.
32. Zhao P, Yao MY, Zhu SJ, et al. Crystal structure of EGFR T790M/C797S/V948R in complex with EAI045. *Biochem Biophys Res Commun*. 2018;502(3):332–7.
33. Wang LX, An YF, Wei XP, et al. In silico screening combined with bioactivity evaluation to identify AMI-1 as a novel anticancer compound by targeting AXL. *Journal of Biomolecular Structure & Dynamics*; 2023.
34. Xie ZL, Ruan W, Guo JJ et al. T5S1607 identified as a antibacterial FtsZ inhibitor: virtual screening combined with bioactivity evaluation for the drug discovery. *Comput Biol Chem*, 2024, 108.
35. Yang Y, Peng RJ, Cheng ZJ et al. Based on virtual screening methods and biological activity assessment, AK-968/13030056 has been identified as a potential mTOR inhibitor. *J Mol Struct*, 2024, 1302.
36. Xu SD, Huang XL, An YF, et al. In silico screening applied in drug discovery: as a novel fourth-generation EGFR inhibitor. *New J Chem*. 2023;47(44):20405–16.
37. Gherghi IC, Girousi ST, Voulgaropoulos AN, et al. Study of interactions between DNA-ethidium bromide (EB) and DNA-acridine orange (AO), in solution, using hanging mercury drop electrode (HMDE). *Talanta*. 2003;61(2):103–12.
38. Sun X, Zhang BL, Luo LX et al. Design, synthesis and pharmacological evaluation of 2-arylurea-1,3,5-triazine derivative (XIN-9): a novel potent dual PI3K/ mTOR inhibitor for cancer therapy (129, 106157, 2022)[J]. *Bioorganic Chemistry*, 2023, 131.

Publisher's note

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