Molecular Diversity (2025) 29:3189–3205
https://doi.org/10.1007/s11030-024-11067-5

ORIGINAL ARTICLE

Identification of STAT3 phosphorylation inhibitors using generative deep learning, virtual screening, molecular dynamics simulations, and biological evaluation for non-small cell lung cancer therapy

Weiji Cai[1,2] · Beier Jiang[3] · Yichen Yin[1,2] · Lei Ma[1,2] · Tao Li[4] · Jing Chen[1,2]

Received: 6 October 2024 / Accepted: 22 November 2024 / Published online: 23 December 2024

Abstract
The development of phosphorylation-suppressing inhibitors targeting Signal Transducer and Activator of Transcription 3 (STAT3) represents a promising therapeutic strategy for non-small cell lung cancer (NSCLC). In this study, a generative model was developed using transfer learning and virtual screening, leveraging a comprehensive dataset of STAT3 inhibitors to explore the chemical space for novel candidates. This approach yielded a chemically diverse library of compounds, which were prioritized through molecular docking and molecular dynamics (MD) simulations. Among the identified candidates, the HG110 molecule demonstrated potent suppression of STAT3 phosphorylation at Tyr705 and inhibited its nuclear translocation in IL6-stimulated H441 cells. Rigorous MD simulations further confirmed the stability and interaction profiles of top candidates within the STAT3 binding site. Notably, HG106 and HG110 exhibited superior binding affinities and stable conformations, with favorable interactions involving key residues in the STAT3 binding pocket, outperforming known inhibitors. These findings underscore the potential of generative deep learning to expedite the discovery of selective STAT3 inhibitors, providing a compelling pathway for advancing NSCLC therapies.

Keywords NSCLC · Drug design · Signal transducer and activator of transcription 3 · Molecular dynamics · Apoptosis

Introduction

Non-small cell lung cancer (NSCLC) is the predominant form of lung cancer [1], constituting approximately 85% of all cases [2]. Although it generally exhibits a slower progression compared to small cell lung cancer (SCLC), NSCLC remains a leading cause of cancer-related mortality worldwide [3, 4]. Traditional treatment modalities, including surgery, chemotherapy [5, 6], and radiotherapy [1], often encounter limitations due to adverse side effects and the development of resistance [3, 7, 8]. In recent years, targeted therapies have emerged as a promising strategy for NSCLC [9], particularly for tumors with specific mutations in genes such as EGFR [10], ALK, and ROS1 [11]. Agents targeting these mutations, including EGFR inhibitors like Gefitinib and ALK inhibitors such as Crizotinib, have shown significant efficacy in patients with these genetic alterations. However, challenges persist, including resistance to targeted therapies and limited options for patients lacking targetable mutations. Ongoing research aims to identify new molecular targets, including MET, RET [12], and KRAS [13], alongside immune checkpoint inhibitors like PD-1/PD-L1 inhibitors [14], which have demonstrated promise in clinical trials and represent an expanding area of therapeutic development for NSCLC.

Signal transducer and activator of transcription 3 (STAT3) has emerged as a critical therapy target for NSCLC, due

Weiji Cai, Beier Jiang have contributed equally to this work.

✉ Tao Li
lit1979@163.com

✉ Jing Chen
chenjing1979@163.com

1 School of Basic Medical Sciences, Ningxia Medical University, 1160 Shengli Road, Yinchuan 750004, Ningxia, China

2 Key Laboratory of Fertility Maintenance Ministry of Education, Ningxia Medical University, Yinchuan 750004, Ningxia, China

3 Navy Medical Research Institute, Naval Medical University, Shanghai 200433, China

4 Department of Oncology, General Hospital of the Ningxia Medical University, Yinchuan 750004, China

to its role in enhancing cancer cell proliferation, survival, and immune evasion [15]. Dysregulated STAT3 signaling is commonly observed in various cancers [16], including NSCLC [17], where it is associated with poor prognosis and resistance to conventional therapies. Therefore, STAT3 has become a pivotal target in the development of novel cancer treatments [18, 19]. Several small-molecule inhibitors with diverse chemical scaffolds have been developed to inhibit STAT3 activity (Fig. 1). For instance, napabucasin, a quinolone derivative, inhibits STAT3-driven transcriptional activity and has shown anticancer effects in clinical trials. Another compound, stattic, selectively inhibits the SH2 domain of STAT3, obstructing its dimerization and activation. Benzimidazole-based inhibitors, such as BP-1-102 [20], effectively bind to the STAT3 SH2 domain, preventing its interaction with phosphotyrosine residues on activated receptors. Additionally, OPB-51602 [21], a novel STAT3 inhibitor featuring a piperazine scaffold, has demonstrated promising anticancer activity in both preclinical and clinical studies. These varied chemical scaffolds highlight the potential of STAT3 inhibitors as viable therapeutic options, particularly for cancers characterized by STAT3 overactivation [22]. Despite these advances, the clinical utility of STAT3 inhibitors has been limited by challenges such as poor bioavailability, off-target effects, and resistance mechanisms. Hence, there is a need for the development of more potent and selective inhibitors that can overcome these barriers and be translated into effective therapies for NSCLC.

Generative deep learning (GDL)-aided drug design is an innovative approach that leverages advanced algorithms to accelerate the discovery of new drug candidates by generating novel chemical structures with desired biological properties [23–25]. In GDL, algorithms such as generative adversarial networks (GANs) [26], recurrent neural networks (RNNs) [27], and variational autoencoders (VAEs) [28] are commonly used to create virtual libraries of potential drug molecules. GDL enables the design of novel compounds by learning from existing datasets of known drugs and predicting new chemical structures optimized for specific biological targets, such as STAT3 [29]. The process begins with the design of a target library, which typically involves selecting molecular scaffolds or fragments that are known to interact with the target protein or pathway of interest [30]. By integrating structure-based design principles [31] and applying GDL, researchers can expand this library to generate novel molecules. Virtual screening follows, where the designed

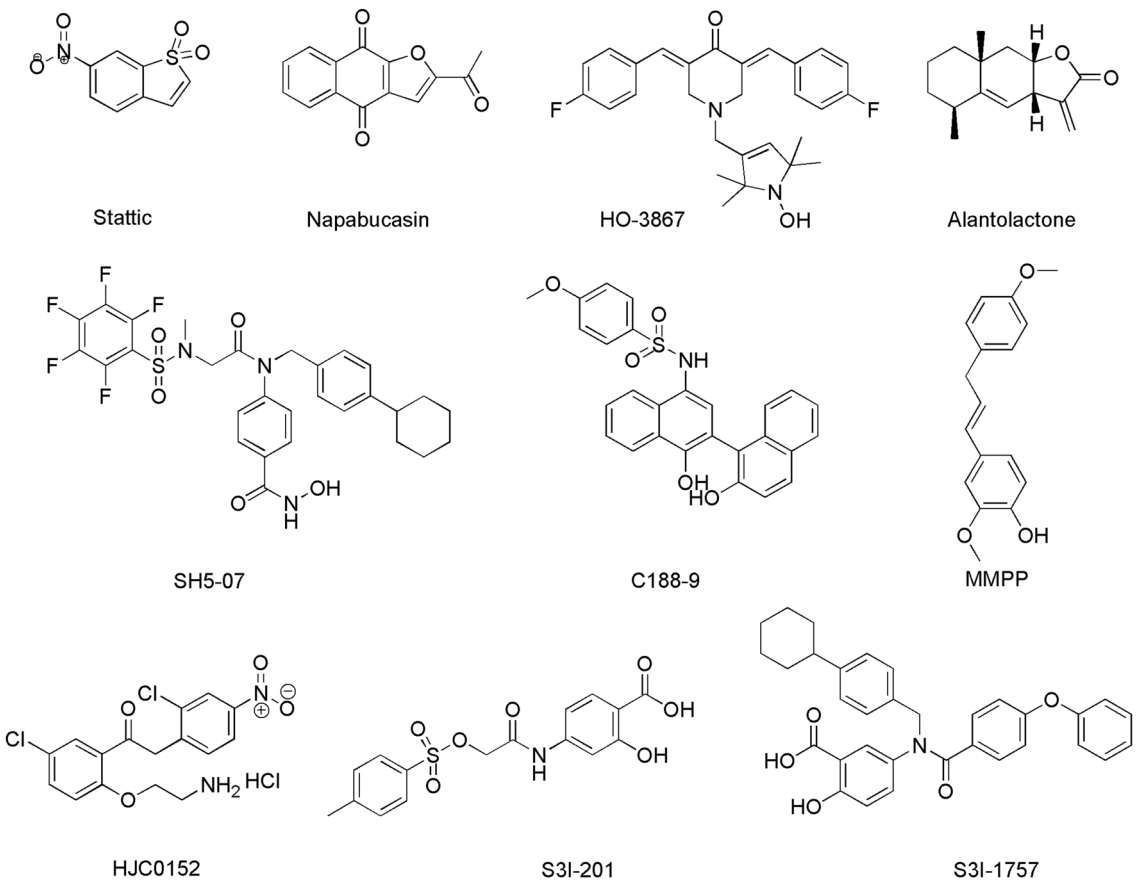

Fig. 1 Chemical structures of STAT3 inhibitors with different scaffolds

molecules are computationally assessed for their ability to bind to the target using techniques like molecular docking or molecular dynamics simulations [32, 33]. By enhancing the chemical diversity and optimizing lead compounds early in the drug discovery process, GDL significantly accelerates the identification of promising therapeutic candidates. This innovative approach has the potential to overcome many of the limitations associated with traditional drug discovery methods and offers a pathway for the development of more effective STAT3 inhibitors for the treatment of NSCLC.

A comprehensive analysis of protein complexes using molecular dynamics and quantum calculations can provide deeper insights into their interactions. In this study, we employed generative deep learning to create customized virtual compound libraries aimed at discovering novel STAT3 inhibitors (Fig. 2). The libraries were generated using a conditional recurrent neural network (cRNN) model [34], which learned distribution patterns to synthesize diverse molecular structures. These compounds underwent a rigorous virtual screening process that included drug-like filtering, molecular docking studies, and synthesis of selected molecules. Ultimately, 95 small molecules were chosen for biological evaluation across H1299 and H441 cell lines. Among these, compounds HG106 and HG110 were subjected to further analysis using colony formation assays and apoptosis assays, alongside characterization of caspase-3-dependent apoptosis biomarkers. Furthermore, we assessed the in vivo inhibition of STAT3 (Tyr705) phosphorylation through immunoblotting and cellular immunofluorescence techniques. Our results demonstrated that HG106 and HG110 exhibited significant efficacy as STAT3 inhibitors, highlighting their potential for therapeutic applications in NSCLC.

Materials and methods

Cell lines and cell culture

NSCLC cell lines NCI-H1299, A549, NCI-H522, and NCI-H441 were obtained from the American Type Culture Collection. MRC-5 and GES-1 cell lines were donated to Prof. Ming-Yao Liu's group at the Shanghai Institute of Key Regulated Life Sciences, East China Normal University, Shanghai, China. NCI-H441, NCI-H522, MRC-5, A549, NCI-H1299, and GES-1 cell lines were cultured in DMEM or RPMI-1640 with 10% fetal bovine serum (Gibco). All cells were cultured in an incubator with 5% CO_2 at a constant temperature of 37 °C.

Generative deep learning to generate virtual molecular library

The generative deep learning model was implanted from previous report. The package was obtained from the Yueshan Li et al. paper [35]. All the parameters were set as default without changing except especial mention in the paper (see Supplementary Table S1). Briefly, the generative model was built to convert molecular SMILES strings into reconstructed SMILES strings using a two-part structure: a feature extractor and a generative layer. The feature extractor

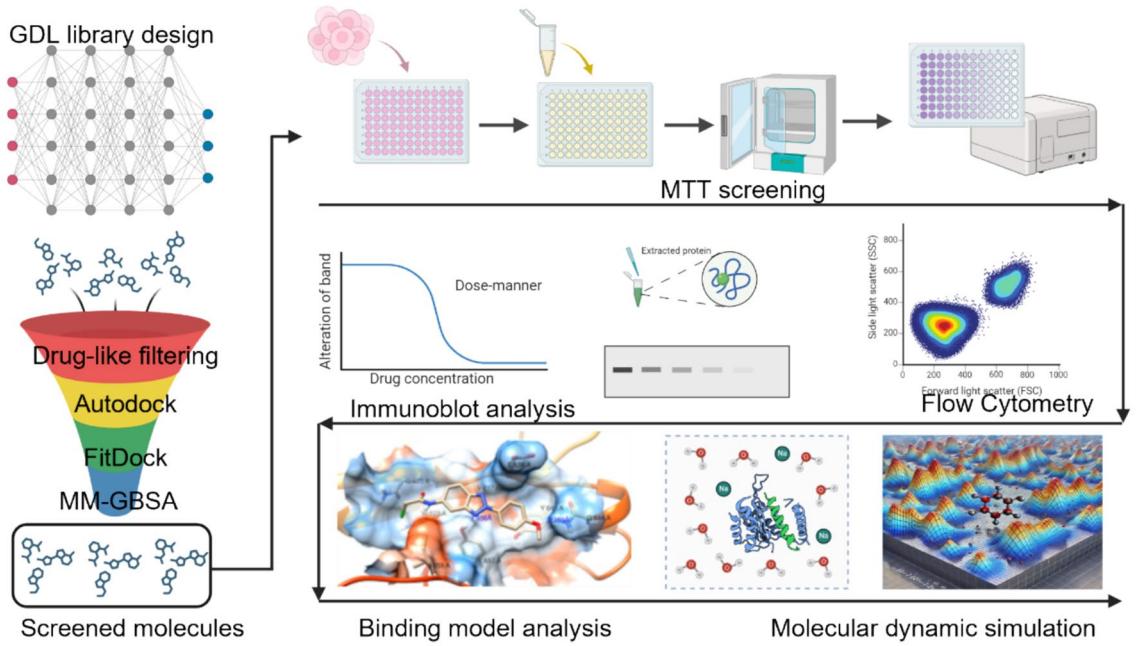

Fig. 2 The flowchart for generative deep learning model, virtual screening, and biological assays to identify merging STAT3 inhibitors

is a one-layer bi-directional LSTM with 512 dimensions, converting input molecules into an initial state vector for the model. This state vector and a one-hot encoded SMILES string feed into a one-layer LSTM with 256 dimensions, followed by a dense layer and a SoftMax activation function. This structure enables a probability distribution over possible grammar production rules, which allows the model to generate chemically valid molecules in each step of SMILES generation. After training, generated molecules were generated with 39,477, and then filtered to remove the duplicate, unstable, and other molecules which cannot transform into 3D molecules ($n = 27{,}863$).

For the pretraining procedure, we collected STAT3 inhibitors from ChEMBL database (https://www.ebi.ac.uk/chembl/, update Sept 1st, 2024), in which IC_{50} is less than 1000 nM, and there were 411 molecules obtained and further transformed as SMILES for the following training. The source data were obtained from ZINC database (https://zinc.docking.org/, update to Sep 1st, 2024) with drug-like pre-set. The molecular structures retrieved from the ZINC database were initially represented in SMILES (Simplified Molecular Input Line Entry System) notation. Subsequently, these SMILES strings underwent a process of standardization and normalization. This involved the elimination of stereochemical details, the removal of counterions associated with salts, and the identification and exclusion of duplicate entries. The aforementioned operations were facilitated by employing the RDKit software package, version 2024.3.2, which is accessible at https://github.com/kuelumbus/rdkit-pypi. As a result of this meticulous curation, a comprehensive dataset comprising approximately 12 million unique molecular entities was assembled for subsequent analysis.

Molecular docking

Molecular docking simulations were performed to predict the binding mode and affinity of the ligand to the target protein STAT3 using AutoDock 4.0. The docking protocol was designed to comprehensively explore the conformational space and provide reliable interaction profiles suitable for a high-impact publication.

Protein Preparation: The three-dimensional structure of the target protein was obtained from the Protein Data Bank (PDB ID: 6nuq), which was reported by Longchuan Bai et al. [36]. The protein structure was prepared using AutoDockTools (ADT) version 1.5.6 following these steps:

1. Removal of Water Molecules and ligand: All crystallographic water and ligand molecules were removed to prevent interference with the ligand binding process.
2. Addition of Hydrogens: Polar hydrogens were added to the protein structure to account for polar hydrogen bonding interactions.
3. Assignment of Kollman Charges: Gasteiger charges were assigned to the protein atoms.
4. Atom Type Assignment: AutoDock atom types were assigned appropriately.

The prepared protein structure was saved in PDBQT format required by AutoDock.

Ligand preparation

The ligand structure was drawn using Kindraw and converted to a three-dimensional conformation using Accelrys ViewerLite 5.0 and further optimizing using openbabel (Version 3.0.0, https://pypi.org/project/openbabel/3.0.0a1/).

Grid preparation

The grid box was centered at the coordinates of the co-crystallized ligand (13.711, 54.024, − 0.083). The grid box size was set to $62 \times 70 \times 70$ points with a grid spacing of 0.375 Å, ensuring sufficient space for the ligand to explore conformational freedom. AutoGrid 4.0 was used to pre-calculate the grid maps for all atom types present in the ligand, including electrostatic and desolvation maps.

Docking parameters: Docking simulations were performed using the Lamarckian Genetic Algorithm (LGA), which combines genetic algorithm search with local optimization, enhancing the efficiency of the docking process.

Number of runs: 100 independent docking runs were conducted to ensure thorough sampling.

Population size: Set to 150 individuals per generation.

Maximum number of evaluations: 2,500,000 energy evaluations per run.

Maximum number of generations: 27,000.

Mutation rate: 0.02.

Crossover rate: 0.8.

Elitism: The top individual was carried over to the next generation.

Local search: Solis and Wets local search method was applied with a probability of 0.06.

Analysis of docking results: The docking results were analyzed using ADT and visualized with UCSF Chimera 1.15. High-resolution images of the protein–ligand complex were generated to illustrate the binding interactions.

Computational Resources: All computations were performed on a workstation equipped with an Intel(R) Xeon(R) CPU E5-2678 processor, 32 GB RAM, and running Ubuntu 20.04 LTS operating system.

Molecular dynamics simulation

Molecular dynamics (MD) simulations were performed using the OpenMM software package (version 7.7) to

investigate the structural dynamics of the STAT 3 protein–ligand complex. The following protocol was employed to ensure a comprehensive and reproducible simulation process [37].

Protein preparation

The crystal structure of the target protein was retrieved from the Protein Data Bank (PDB ID: 6nuq). Missing residues and side chains were modeled using PDBFixer and pdb2qpr package, and protonation states were assigned appropriate for pH 7.4 using the PropKa algorithm. Hydrogen atoms were added using the pdb2pqr module in OpenMM.

Ligand preparation

The ligand structure was obtained by preparing using Kindraw software (HG106 and HG110). Geometry optimization and partial charge assignment were performed using the AM1-BCC method implemented in the OpenEye Quacpac toolkit 666. Ligand force field parameters were generated using the LigParGen (https://zarbi.chem.yale.edu/ligpargen/, CM1A with neutral charge). The PDB file of ligand and relative force field were downloaded and prepared for the next step.

System assembly

The prepared protein and ligand were combined to form the protein–ligand complex using ChimeraX. The complex was then solvated in an explicit water box using the TIP3P water model, with a padding distance of 10 Å from the protein surface to the box edges. Sodium and chloride ions were added to neutralize the system and achieve a physiological ionic strength of 0.15 M using the Modeller module of openmm.

Force field parameters

The AMBER99SB force field was employed for the protein, while the ligand was parameterized with GAFF2 parameters. The system's topology and parameters were generated using the ForceField class in OpenMM, integrating both protein and ligand parameters seamlessly.

Energy minimization

Energy minimization was conducted to relieve any steric clashes and optimize the initial geometry of the system. The conjugate gradient algorithm in OpenMM was used with a maximum of 5,000 steps or until the energy gradient converged below 0.01 kJ/mol/Å.

Production MD simulation

An unrestrained production MD simulation was executed for 200 ns under NPT conditions at 300 K and 1 atm. The Langevin integrator was used with a time step of 4 fs. All covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm, allowing for the larger integration time step. Long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method with a real-space cutoff of 10 Å. van der Waals interactions were smoothly switched off between 8.0 Å and 10.0 Å to reduce cutoff artifacts.

Data collection

Coordinates were saved every 20 ps, resulting in 10,000 frames for analysis. The simulation was monitored by recording the energy components, temperature, pressure, and volume at regular intervals to ensure system stability.

Software and hardware

All simulations were performed using OpenMM on a workstation equipped with an NVIDIA V100 GPU 16 GB, leveraging GPU acceleration for efficient computation.

Calculations of molecular properties

The physicochemical properties of the obtained molecules were computed using RDKit (Version 2024.3.2, https://github.com/kuelumbus/rdkit-pypi), a cheminformatics toolkit. A set of critical molecular descriptors were calculated to evaluate the drug-likeness and other relevant properties of the compounds. The following properties were assessed:

Molecular weight (MW)

The molecular weight of each compound was calculated as the sum of the atomic weights of all atoms in the molecule, including isotopes and charge contributions, to determine the overall size of the molecule.

Water–octanol partition coefficient (LogP)

The LogP value was computed using the Crippen method to estimate the hydrophobicity of the compound. This coefficient indicates the compound's ability to partition between the water phase and the octanol phase, which

correlates with its lipophilicity and drug absorption potential.

Qualitative estimate of drug-likeness (QED)

The drug-likeness of each compound was evaluated using the Quantitative Estimate of Drug-likeness (QED), which ranges from 0 to 1. This descriptor combines several molecular properties, including molecular weight, rotatable bonds, hydrogen bond donors/acceptors, and more, to predict how "drug-like" a compound is based on its physicochemical features.

Topological polar surface area (TPSA)

TPSA was calculated to estimate the polar surface area of the molecule, which plays a significant role in predicting intestinal absorption, blood–brain barrier penetration, and overall bioavailability. This descriptor helps to determine the molecule's ability to cross biological membranes.

Water solubility (LogS)

The water solubility (LogS) was predicted using the Ali method, which correlates molecular features with experimental solubility data. LogS is an important factor influencing drug absorption, distribution, and excretion.

Hydrogen bond donors (HBD)

The number of hydrogen bond donors was calculated by counting the number of hydrogen atoms attached to electronegative atoms (like oxygen or nitrogen) that can participate in hydrogen bonding interactions.

Hydrogen bond acceptors (HBA)

The number of hydrogen bond acceptors was determined by counting the number of electronegative atoms (oxygen, nitrogen, etc.) that have lone pairs of electrons capable of accepting a hydrogen bond from donors.

To ensure accurate predictions, all compounds were first converted into standard SMILES (Simplified Molecular Input Line Entry System) strings using RDKit, followed by the calculation of the above properties. Each descriptor was obtained by running the appropriate RDKit functions that are designed to handle different molecular types, such as aromatic rings, heteroatoms, and functional groups.

Uniform manifold approximation and projection (UMAP)

To elucidate the relational similarities among the source data, target data, and generated data, we employed UMAP (umap-learn 0.4.6, https://pypi.org/project/umap-learn/) to create two-dimensional projections of high-dimensional data distributions. This analysis involved 3,000, 1,000, and 2,000 randomly selected molecules from the source, target, and generated datasets, respectively. All the images were generated using Python = 3.8 with scipy package (Version 1.10.1).

Tree maps (TMAPs)

Scaffold and fingerprint diversity were assessed and visualized utilizing the Platform for Tree maps (tmap, https://tmap.gdb.tools/), based on a random selection of 15,678 molecules from the source data, target data, generated data, and filtered molecules, respectively.

Cell viability assay

Cells were collected from culture dishes, counted, resuspended, and seeded into 96-well plates. After allowing them to adhere overnight, we replaced the medium with fresh medium containing various concentrations of the drug and incubated the cells for an additional 72 h. To fix the cells, we added 25 µL of 50% trichloroacetic acid (Sangon Biotech, China) to each well and refrigerated the plates at 4 °C overnight. Following fixation, the plates were washed and dried before adding 50 µL of 0.4% sulforhodamine B solution (A426137, Sangon Biotech, China) to each well for staining. After a 10-min incubation, excess dye was removed by washing the plates several times with 1% glacial acetic acid (A501931, Sangon Biotech, China) until the wash solution was clear. The plates were then air-dried, and 100 µL of 10 mM Tris buffer (T8060, Solarbio, China) was added to each well to dissolve the dye. Absorbance was measured at 515 nm using a plate reader.

Colony-forming assay

Non-small cell lung cancer cell lines NCI-H441, NCI-H1299, NCI-H522, and A549 were seeded into 6-well plates at densities ranging from 1,000 to 3,000 cells per well and cultured for 24 h to allow for cell attachment. The following day, the cells were treated with HG106 and HG110 at given concentrations of 0.5 µM and 1 µM, respectively, for a period of 7 to 10 days, with three replicate wells for each group. The culture medium was refreshed and the drugs were re-applied every three days. After the treatment period, cells were fixed with 4% paraformaldehyde (P1110, Solarbio, China) for 30 min. The plates were washed three times with running water and air-dried. Cells were then stained with 0.1% crystal violet (A600331, Sangon Biotech, China), washed with ultrapure water, dried, and photographed. Colonies containing more than 50 cells were counted.

Soft-agar colony formation assay

A 1.2% agar solution (A8190, Solarbio, China) was combined with growth medium and poured into the wells of a 6-well plate at a volume of 2 mL per well. The mixture was allowed to solidify at room temperature. Once the first agar layer had set, a second layer containing 0.5% agar and 2000 cells per well were added, using 1 mL per well. Fresh medium was then added on top of the second layer, and the cells were incubated for approximately 4 weeks. During this period, the medium supplemented with the specified drugs was refreshed every 3 days. At the end of the incubation period, the colonies were photographed for further analysis.

Western blotting analysis

Cells were seeded into 60 mm culture dishes and allowed to adhere. After adhesion, different concentrations of the target compound drugs were added to the medium, and the cells were incubated for an additional 24 h. The medium was then removed, and the cells were washed three times with pre-chilled 1 × PBS (P1010, Solarbio, China). Next, 150 μL of cell lysis buffer (HY-K1001, MCE, USA) was added to each dish, and the protein lysate was collected into EP tubes (615,601, NEST, China) using a spatula. The lysates were kept on ice for 30 min, with gentle shaking for 15–20 s every 10 min to ensure complete lysis. Protein concentrations were measured using a BCA assay kit (23,235, Thermo, USA), and each sample was adjusted to approximately 50 μg of protein. The protein samples were mixed with 5X protein loading buffer (C516031, Sangon Biotech, China) and heated at 100 °C for 10 min to denature the proteins. The denatured proteins were then separated by electrophoresis until the bromophenol blue dye reached about 1 cm from the bottom of the gel. After electrophoresis, the proteins were transferred to a membrane, which was then incubated with the primary antibody overnight at 4 °C on a shaker. The following day, the membrane was incubated with a secondary antibody diluted in skimmed milk for 1 h on a shaker to complete the detection process.

The membrane was visualized using an Odyssey infrared two-color laser scanning imaging system. Reagents and antibodies: STAT3 (cat. no. 9139), antibodies against phosphorylated (p)-STAT3 (Tyr705; 9145), Bcl-2 (4223), Caspase-3 (4691), Bax (Ser465; 467) were purchased from Cell Signaling Technology. GAPDH and β-actin (cat. no. M1210-2) were purchased from Hangzhou Huaan Biotechnology. Other reagents, such as dimethyl sulfoxide, were obtained from Sigma-Aldrich (Merck KGaA).

Apoptotic assay

NCI-H441 cells were treated with the specified drugs in 6-well plates for 24 h. After treatment, the cells were detached using EDTA-free trypsin (T1350, Solarbio, China) and collected into labeled EP tubes. Following centrifugation, the supernatant was removed. The cell pellet was resuspended gently in 500 μL of Binding Buffer (KGA1102, KeyGEN, China) to create a single-cell suspension. Next, 5 μL of Annexin V-FITC (KGA1102, KeyGEN, China) was added to each sample and mixed thoroughly, followed by the addition of 5 μL of propidium iodide (KGA1102, KeyGEN, China). The samples were incubated at room temperature in the dark for 5 to 10 min. Flow cytometry (BD Accuri™ C6 Plus, USA) analysis was then performed, examining 20,000 cells per sample. Early apoptotic cells were identified as those positive for Annexin V-FITC but negative for propidium iodide, while late apoptotic cells were positive for both Annexin V-FITC and propidium iodide. The results were expressed as the percentage of each apoptotic population relative to the total number of cells analyzed (FlowJo 10.8.1).

Immunofluorescence

Cell slides were first soaked in 75% ethanol for 1 h and then dried by holding them with forceps approximately 2–3 cm from an alcohol lamp. The prepared slides were placed into a 24-well plate, and cells were seeded onto the slides and incubated until they adhered to the surface. After cell adhesion, the medium was replaced with fresh medium containing varying concentrations of the target compounds, and the cells were incubated for an additional 24 h. Subsequently, the cells were treated with interleukin-6 (PeproTech, USA)) for 2 h. The cells were then fixed on ice using 4% paraformaldehyde (P1110, Solarbio, China) for 20 min. Following fixation, the cells were permeabilized with 0.3% Triton X-100 (T8200, Solarbio, China) for 30 min and washed three times with PBS, each wash lasting 10 min. Blocking was performed using 1% bovine serum albumin (A8020, Solarbio, China) for 1 h to prevent non-specific antibody binding. After blocking, 100 μL of diluted primary antibody was added to each slide and incubated overnight at 4 °C on a shaker. The next day, the slides were washed with 1 × PBS for 10 min to remove any unbound antibody.

DFT calculation

The optimization of the compounds was performed by using PySCF. For this process, B3LYP/6-311G + (d,p) level basic set of DFT was used. Afterward, the Frontier Molecular Orbitals (FMO) which were the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital

(LUMO) were extracted and the energy gap was calculated by the difference between HOMO and LUMO energy values. Furthermore, quantum chemical parameters were calculated in order to predict the behavior of compounds and to get information about the transition state energies.

Virtual screening strategies

Virtual screening was performed to identify potential small-molecule inhibitors for the target protein using a combination of drug-like filtering, molecular docking, and binding affinity calculations. Initially, a compound library was screened based on revised Lipinski's Rule of Five (Molecular weight ≤ 700 Da; LogP ≤ 5; Hydrogen bond donors ≤ 5; Hydrogen bond acceptors ≤ 10) to select drug-like candidates with favorable pharmacokinetic properties.

To optimize the enrichment of active compounds, a hierarchical two-step pipeline was implemented for the structure-based virtual screening of potential inhibitors. Initially, AutoDock Vina was employed to dock a large compound library into the target protein's binding site, which allowed for the rapid elimination of inactive or poorly binding compounds. The compounds were ranked according to their predicted binding affinities, and the top 10% were selected for further analysis.

In the second step, an in-house developed scoring function, CYSCORE [38], was used to refine the selection of compounds based on their predicted binding energies and other physicochemical properties. The top 10% of compounds from the AutoDock Vina docking were further analyzed using FitDock [39], a compound alignment and docking program. This tool assessed the structural similarities between the compounds, clustering them into groups based on common chemical scaffolds and binding modes. This two-step virtual screening strategy allowed for the identification of high-affinity inhibitors of MDM2, streamlining the selection of promising candidates for further experimental validation.

Additionally, Molecular Mechanics, General Born Surface Area (MM/GBSA) calculations (gmx_MMPBSA package with the GROMACS, 100 ps) were performed to refine the binding affinity predictions and assess the stability of the ligand–protein complexes. The final selection of top-ranked molecules was based on their docking scores, ligand efficiency, drug-likeness, with compounds that demonstrated favorable results proceeding to experimental validation.

Statistical analysis

All experiments in this study were repeated three times independently, and all data were analyzed using GraphPad Prism 9.0. One-way analysis of variance (ANOVA) was used to compare the drug treatment group, positive control group, and blank control group. The results were expressed as mean ± standard deviation (SD). Differences were considered statistically significant at $P < 0.05$.

Results

Generation of molecular libraries using generative deep learning

We employed a generative deep learning approach to construct a diverse library of molecular structures specifically designed to target the active site of STAT3 (Fig. 3A). Initially, we curated a dataset of STAT3 inhibitors exhibiting high biological activity ($IC_{50} < 1000$ nM) from the ChEMBL database (n = 411) as the target dataset. The AlogP value of STAT3 was selected as standard criteria for the subsequential and crude screening (Supplementary Fig. S1). Additionally, approximately 12.0 million compounds from the ZINC database were utilized as the source dataset, providing a broad spectrum of pharmacophoric fragments for model training (Supplementary Fig. S2). Using these datasets, we generated a library of 15,678 unique molecular structures, with any duplicates meticulously removed to ensure novelty.

The generated molecules were visualized using Uniform Manifold Approximation and Projection (UMAP) to assess their distribution in the chemical space relative to both the source and target datasets. UMAP is a dimensionality reduction technique that visualizes high-dimensional data by preserving both local and global structures, enabling a more interpretable representation of complex data distributions [40]. It was used in this study to assess the distribution of generated molecules, as it effectively reduces the high-dimensional molecular feature space into two dimensions, allowing for a clear visualization of the relationships between source, target, and generated data, and providing insights into the diversity and similarity of the molecular structures [41]. As shown in Fig. 3B, the generated molecules (represented by gray dots) demonstrate a clear transition from the chemical space of the source data (red dots) towards that of the target data (blue dots). This distribution indicates that the generative model effectively captured and transferred the relevant features from the source data to the target dataset, creating novel molecular entities aligned with the desired chemical properties of STAT3 inhibitors.

Further analysis of the physicochemical properties of the generated molecules, presented in Fig. 3C, revealed that 98.3% of the molecular scaffolds in the generated library were structurally distinct from those in both the source and target datasets. This substantial scaffold diversity suggests the successful generation of novel chemical space, providing a robust foundation for subsequent molecular docking

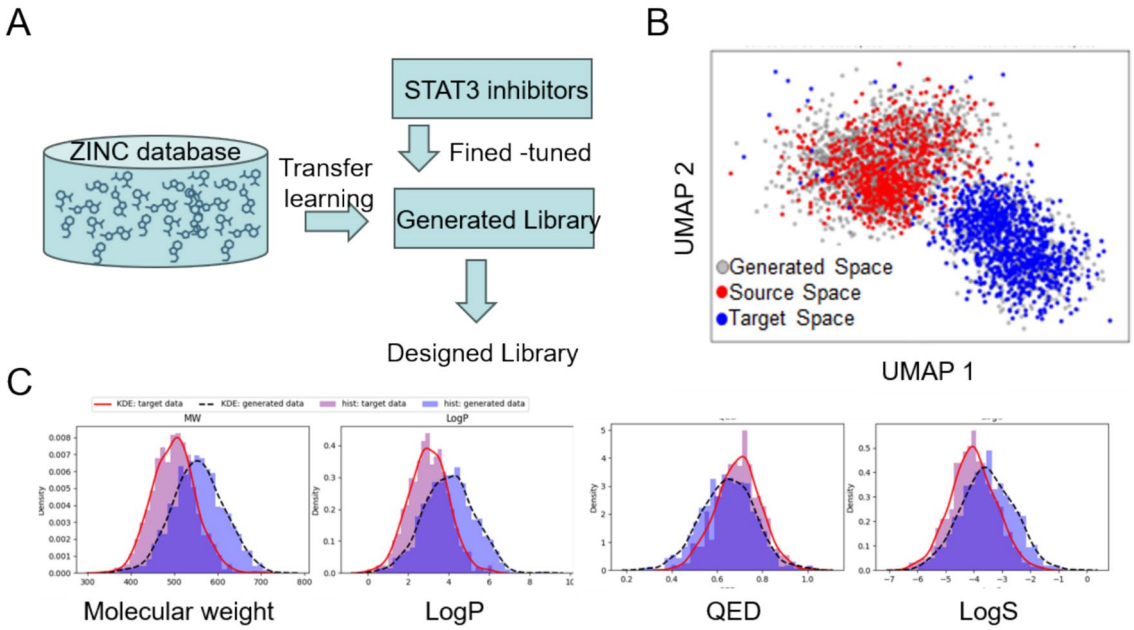

Fig. 3 Generation of a virtual compound library against STAT3 using the cRNN-ased generative model. **A** schematic illustration of library using GDL methods. **B** chemical space of generated space (gray dot), source space (red dot), and target space (blue dot) by transfer learning. **C** the distributions of molecular weight, logP, QED, and logS of targeted data and the generated data by GDL generation methods

and virtual screening efforts aimed at identifying promising STAT3 inhibitors.

Virtual screening and molecular docking

After generation the library, we then conducted the virtual screening to obtain the potential STAT3 inhibitors (Fig. 2) with STAT3 protein structure (Supplementary Fig. S3) with optimized method (Supplementary Fig. S4). The screened molecules were presented using tree maps (TMAPs) according to the following features, including RECP-based structural similarity, molecular properties, and docking scores (Supplementary Fig. S5). These TMAPs well presented the unsupervised visualization of high-dimensional data with creating a 2D layout of spanning tree in the chemical space. From the top-ranked 95 molecules (Supplementary Fig. S6), we selected the molecules from the current library containing benzotriazole scaffold, and further conducted the biological screening to identify higher anticancer activity.

HG106 and HG110 inhibits the viability of NSCLC cells

Following the outcomes of the virtual screening, we selected 90 novel small-molecule compounds for synthesis. These synthesized compounds were subsequently evaluated for their cytotoxic effects against NSCLC cell lines NCI-H441 and NCI-H1299. Among them, two compounds, i.e., HG106 and HG110, exhibited a significantly enhanced inhibitory effect on the proliferation of both H441 and H1299 cells (Supplementary Fig. S7A), thus identifying HG106 and HG110 as a lead candidate for further investigations. Subsequently, HG106 and HG110 were tested on a panel of four NSCLC cell lines (NCI-H441, NCI-H1299, NCI-H522, A549) and two normal cell lines (GES-1, MRC-5) over a 72-h incubation, and IC_{50} values were determined using the SRB assay. The results demonstrated that HG106 and HG110 markedly suppressed the viability of A549, NCI-H522, NCI-H441, and NCI-H1299 cells (Supplementary Fig. S7B). Furthermore, HG106 and HG110 exhibited minimal cytotoxicity towards normal human cell lines, including MRC-5 fibroblasts and GES-1 gastric epithelial cells, indicating a favorable safety profile.

HG106 and HG110 inhibits the proliferation of NSCLC cells

HG106 and HG110 can efficiently suppress proliferation of lung cancer cells (A549, H522 H441, and H1299) compared with normal cells (MRC-5 and GES-1). To further evaluate cancer cell growth, colony formation assays was conducted. As given in Fig. 4A, NSCLC cell lines NCI-H441, NCI-H1299, NCI-H522, and A549 were dosage-dependent (0.5 and 1.0 μM) exposed to HG106 and HG110 for 7 to 10 days. The results demonstrated that HG106 and HG110 significantly suppressed colony formation in

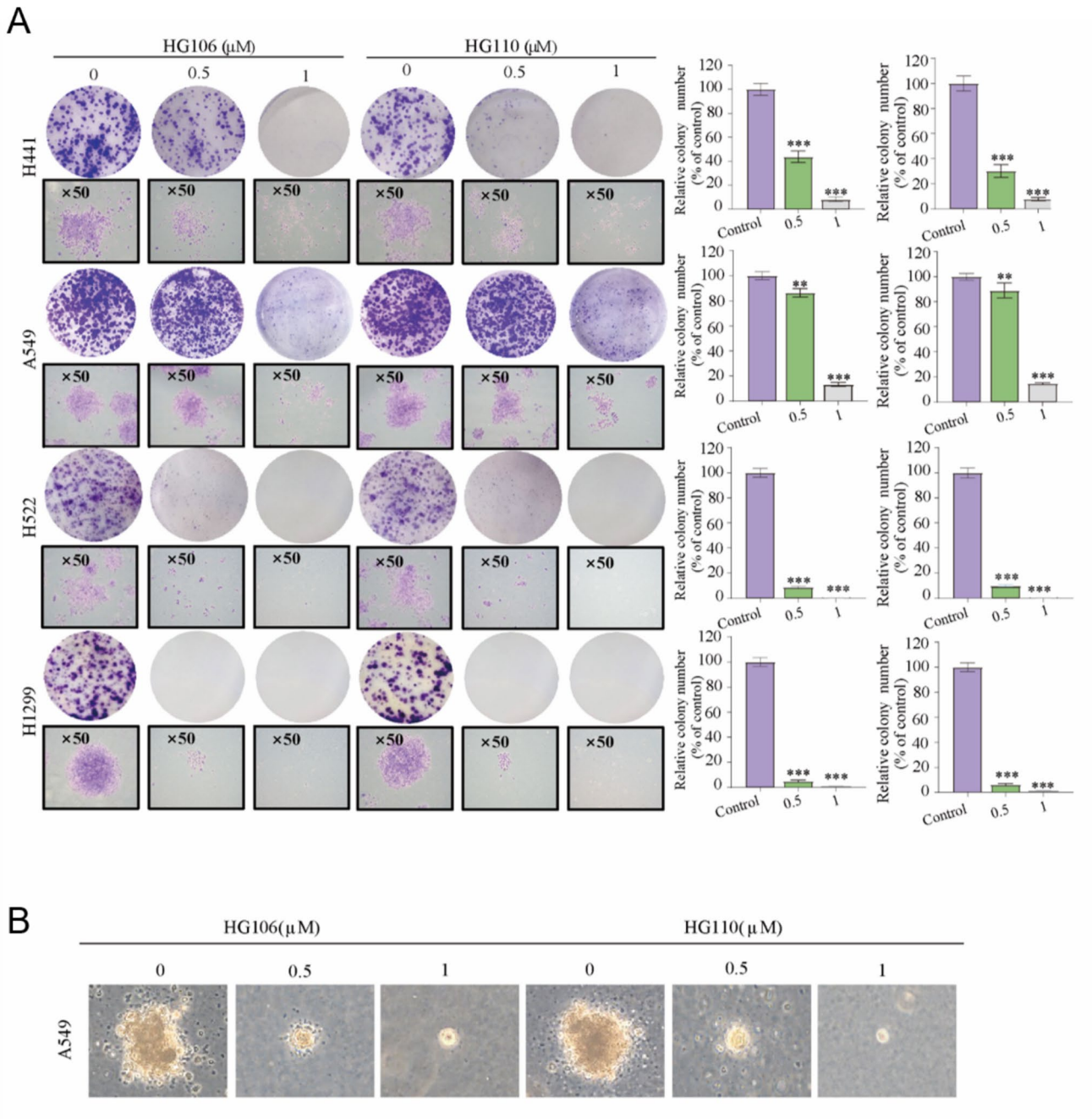

Fig. 4 HG106 and HG110 efficiently suppress cancer cell growth. **A** representative image (left panels) and quantification (right panels) of concentration-manner colony formation assays of cancer cells treating with HG106 and HG110, respectively. **B** bright-field images of soft agar colony formation of A549 cells treated with HG106 and HG110, respectively. All the data were presented as Mean±S.D. Statistical analysis was conducted by one-way ANOVA (***$p < 0.001$)

all tested NSCLC cell lines at a concentration of 1 μM Fig. 4A. This inhibitory effect was further corroborated by a soft agar colony formation assay, which yielded consistent results (Fig. 4B). Collectively, these findings indicate that HG106 and HG110 effectively inhibits the proliferation of NSCLC cells in vitro.

HG106 and HG110 promotes apoptosis in NSCLC cells

Flow cytometric analysis revealed that HG106 and HG110 significantly enhanced apoptotic cell death in H441 cells compared to the control group (Fig. 5A). This pro-apoptotic

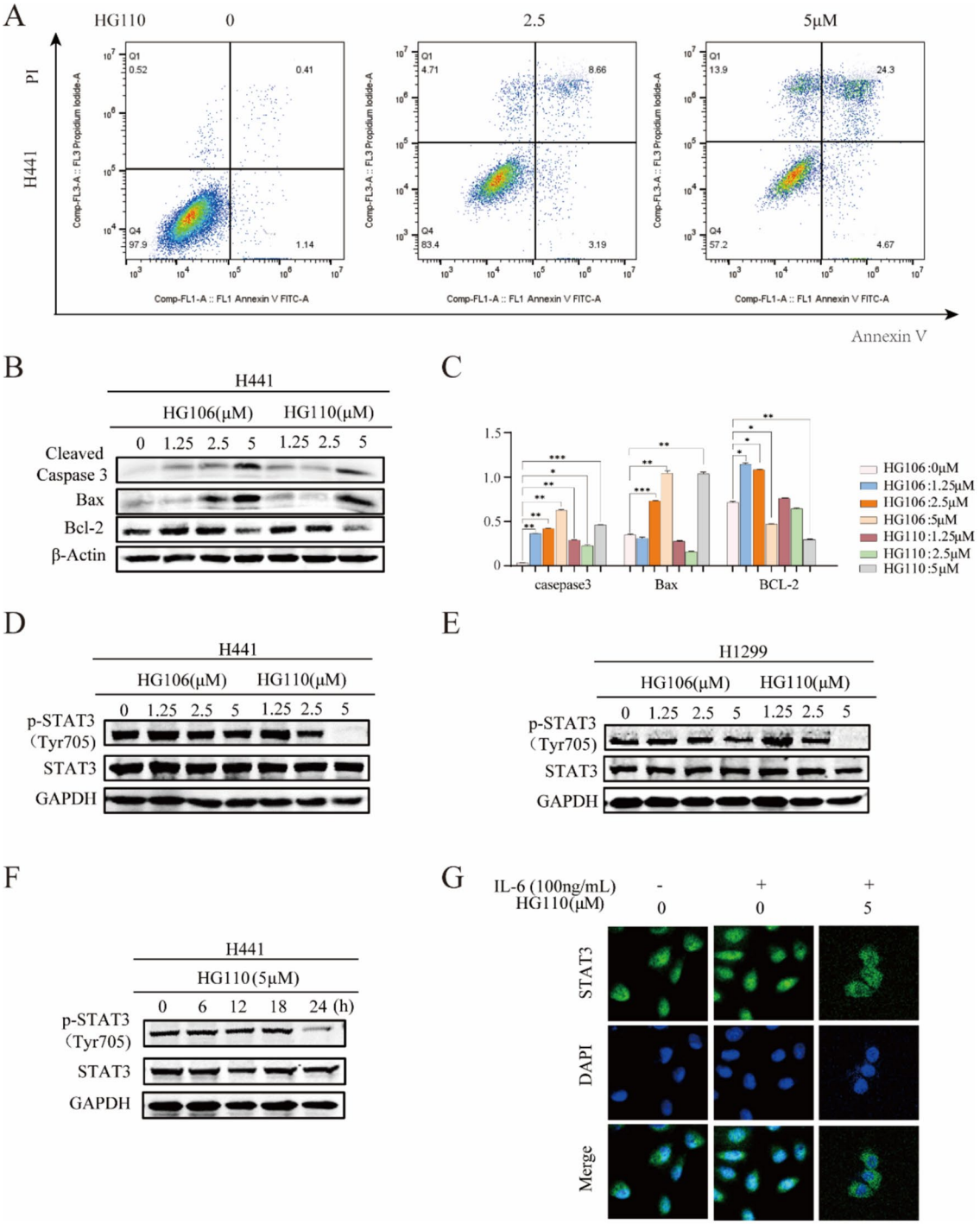

Fig. 5 **A** The percentage of apoptotic cells quantitatively measured by using flow cytometry for the H441 cells stained by Annexin V/PI. Cells were treated with HG106 and HG110 at concentration of 5 uM for 24 h, respectively. DMSO was selected as control. **B** Immunoblot bands and quantitative measurement revealed apoptosis-associated protein expression levels after dose-dependent treatment against H441 cells. **C** Immunoblot analysis reveals the effect of HG110 on the phosphorylation levels of STAT3 at Tyr705 compared with HG106 at 24 h in H411 and H1299 cells. GAPDH was selected as inner standard. **D** HG110 suppressed phosphorylation levels of STAT3 at 24 h. **E** Immunofluorescent staining revealed the impacts of HG110 on IL6-promoted nucleus translocation of STAT3 in H441 cells. Scale bar is XX um. All the data were presented as Mean±S.D. Statistical analysis was conducted by one-way ANOVA (*$p<0.05$, **$p<0.01$, ***$p<0.001$)

effect was further substantiated by western blot analysis, which demonstrated an upregulation of the pro-apoptotic protein Bax and the effector protein caspase-3, alongside a concurrent downregulation of the anti-apoptotic protein Bcl-2 following HG106 and HG110 treatment (Fig. 5B, C). These findings suggest that HG106 and HG110 induce apoptosis in H441 cells through the modulation of key apoptotic regulators.

HG106 and HG110 exert anti-tumor activity by suppressing the activation of STAT3

Western blot analysis was employed to evaluate the inhibitory effect of HG106 and HG110 on STAT3 activation in H441 and H1299 cells following treatment with various concentrations of HG106 or HG110 for 24 h. The results indicated that HG106 and HG110 at a concentration of 5 μM significantly suppressed the phosphorylation of STAT3 (Fig. 5D–F). Additionally, immunofluorescence staining demonstrated that HG106 and HG110 treatment led to a notable reduction in nuclear STAT3 levels, suggesting an inhibition of STAT3 nuclear translocation (Fig. 5G).

Binding model analysis and molecular dynamics simulations

As above-mentioned analysis, HG106 and HG110 presented the excellent anti-proliferation and suppression phosphorylation of STAT3 in vitro. In order to analyze the binding model between STAT3 and HG106 or HG110, we obtained the potential binding conformation of STAT3-HG106 and STAT3-HG110 by molecular docking.

As given in Fig. 6A, HG106 molecules well docked into the STAT3 bioactive pocket, which was widely used to design STAT3 inhibitors [36], and docking score was − 8.5 kcal/mol. By presenting using LigPlot software, we can observe that residue of STAT3, including Trp623, Gln635, Ser636, Val637, Tyr640, Tyr657, and Ile659 were interacted with HG106 by the hydrophilic interaction. Moreover, the amino group of Glu638 interacted with nitrogen atom of triazole ring by hydrogen bond, and NH2 of Gln644 amide group interacted with oxygen of methoxyl of HG106 by hydrogen bond. To further analyze the structure of STAT3-HG106, molecular dynamic simulation was conducted to evaluate the stabilization in 200 ns. Among the 200 ns simulation, system fast achieved the equilibrium at around 10 ns (Supplementary Fig. S8A), and HG106 molecule can stabilize to bind with the pocket of STAT3. To determine the fluctuating residues, we analyzed the RMSD of residue of residues among the 200 ns simulation, and alignment between RMSD and RMSF of residues (Fig. 6B) clearly revealed that the HG106 molecules can strongly interact with the residue of active pockets. The profile of hydrogen bonds altered during the simulation, and the hydrogen bond (Fig. 6C) between methoxyl group and Gln644 was stable.

The free energy landscape was conducted to determine the collective atomic motions during the simulation process with principal component analysis (PCA). As given in Supplementary Fig. S8B, C, the dark blue indicated the minimum valley of system energy, while the difference between top point and lower point is only 1.35 kcal/mol. Additionally, there are three minimum points during the simulation process, and difference is very small. These results may be caused by the weak hydrogen bonds between STAT3 and HG106, consistent with fluctuations of hydrogen bonds (Fig. 6C).

Similarly, HG110 molecules also well docked into the STAT3 bioactive pocket consistent with HG106. By analysis of binding model (Fig. 6D), we can observe that residue of STAT3 including Gln635, Trp623, Val637, Glu638, Pro639, Tyr640, Ile659, and Ser636 were interacted with HG110 by the hydrophilic interaction. Additionally, the NH2 of Gln644 amide part strongly interacted with two methoxyl groups of HG110 by forming hydrogen bonds, while hydroxyl of Tyr657 interacted with nitrogen atom of triazole ring by hydrogen bond. Among the 200 ns simulation (Supplementary Fig. S9A), we can observe that the RMSD of ligand (Fig. 6E) was more stable compared with HG106, and the comparison between RMSD and RMSF of residues also validated that HG110 molecule strongly interacted with STAT3 active pocket (Fig. 6E). Compared with HG106, hydrogen bonds (Gln644 and Tyr657) between HG110 and STAT3 were far more and stable than HG106, implying that the HG110 molecule was more strongly interacted with STAT3 compared with HG106, which is consistent with the above-mentioned biological activities (Fig. 5C). Moreover, the free energy landscape of STAT3-HG110 (Supplementary Fig. S9B, C), only one minimum point was observed compared with STAT3-HG106 complex (Supplementary Fig. S8B, C). This result also validated that stronger hydrogen bonds can stabilize the protein–ligand complex.

By further analysis of RMSD of HG106 and HG110 molecules during the 200 ns simulation, we observed that average RMSD value of HG106 and HG110 were 0.19 Å and 0.16 Å, respectively. This lower average RMSD value of these ligands indicated higher stability. Subsequentially, we also conducted the quantum calculation to obtain the frontier molecular orbitals of HG106 and HG110. The energy gap between the LUMO and HOMO is often used to determine the kinetic stability. As shown in Supplementary Fig. S10A, B, the energy gap of HG110 (9.6300 eV) was higher than HG106 (9.4757 eV), implying that lower reactivity of HG110 and suppression of STAT3 may be originated from the stronger binding, not irreversible covalent binding.

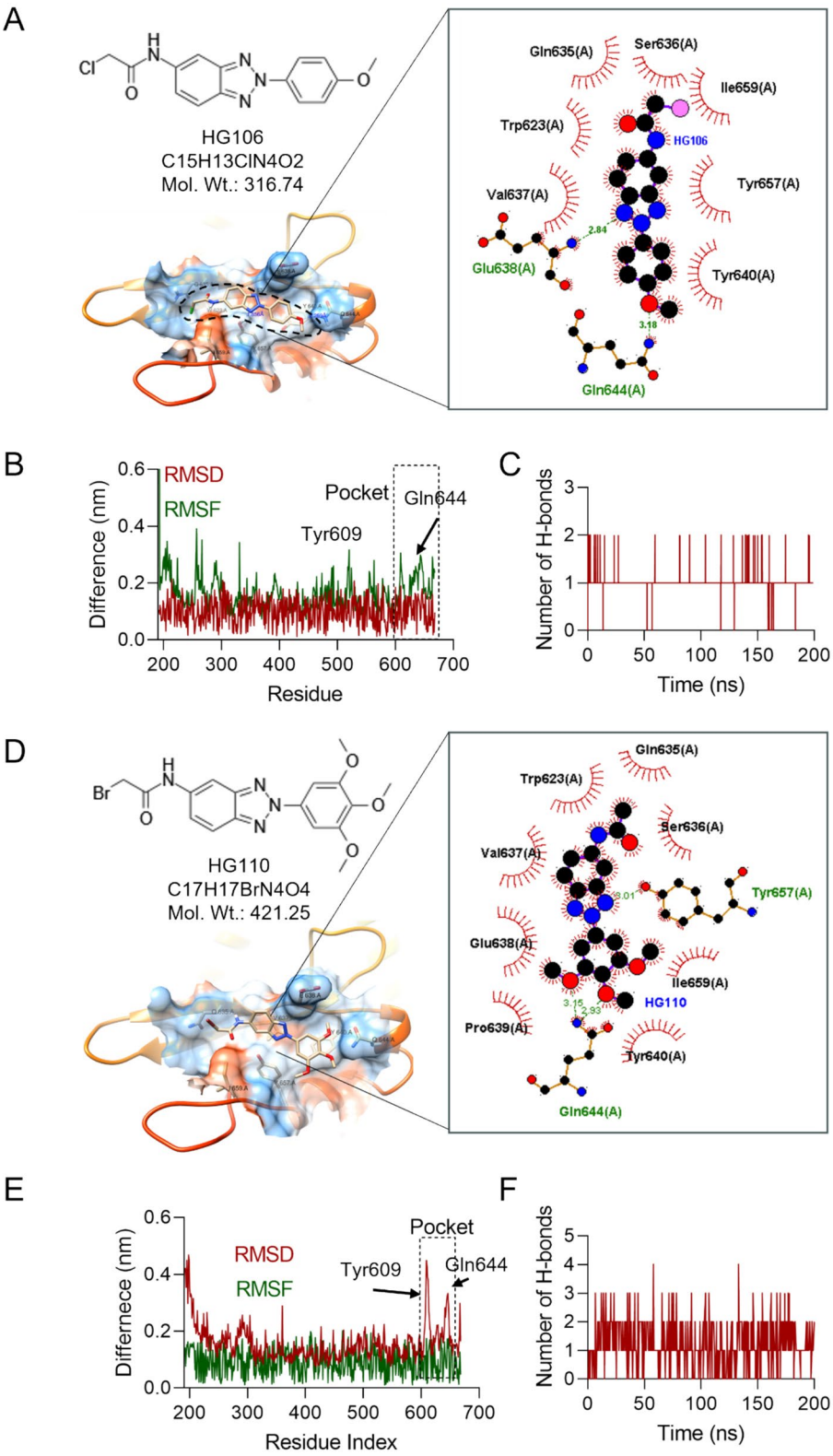

Fig. 6 Molecular dynamic simulation to reveal the binding model of HG106 and HG110 with STAT3. **A** Molecular structure, binding model, and planar residue–ligand interaction of HG106. **B** Analysis of residue root mean square deviation (RMSD) in compound HG106 and STAT3 complex for protein (STAT3) stability. **C** Profile of hydrogen bonds between HG106 and STAT3 among molecular dynamic simulation. **D** Molecular structure, binding model, and planar residue–ligand interaction of HG110. **E** Analysis of residue root mean square deviation (RMSD) in compound HG110 and STAT3 complex for protein (STAT3) stability. **F** Profile of hydrogen bonds between HG110 and STAT3 among molecular dynamic simulation

Discussion

The aberrant activation of STAT3 in vivo plays a pivotal role in driving cancer cell proliferation and regulating multiple tumor-associated pathways, including cell cycle [36], apoptosis resistance [42], angiogenesis [43], and immune evasion [44]. Targeting the overactivation of STAT3, particularly its phosphorylation at Tyr705, is therefore a promising

therapeutic strategy for cancer treatment. In this study, we employed a generative deep learning model to design and synthesize a library of potential STAT3 inhibitors. From this library, two lead compounds, HG106 and HG110, were identified, both derived from a benzotriazole scaffold. Notably, HG110 demonstrated superior efficacy compared to HG106 in inhibiting STAT3 phosphorylation at Tyr705, leading to enhanced induction of apoptosis in NSCLC cells.

In vivo experiments further confirmed the therapeutic potential of HG110, showing its ability to effectively inhibit IL6-induced translocation of phosphorylated STAT3 (p-STAT3) to the nucleus, thereby disrupting STAT3-mediated transcriptional activity. Upon IL6 binding to its receptor, gp130, JAK1 and JAK2 kinases are activated, leading to phosphorylation of STAT3 at Tyr705 [45, 46]. This phosphorylation triggers the dimerization of STAT3, which subsequently translocates to the nucleus to initiate the transcription of genes involved in cell survival, proliferation, and immune evasion [47]. Our studies indicate that HG110 effectively inhibits this phosphorylation step, thereby preventing the nuclear translocation of STAT3 and its downstream transcriptional activity. Molecular dynamics simulations provided additional insights, revealing that HG110 forms stable interactions with key residues in the active pocket of STAT3, thus stabilizing the STAT3-HG110 complex and enhancing its inhibitory effect. These findings collectively highlight HG110 as a potent and selective STAT3 inhibitor with the capability to suppress STAT3 phosphorylation and promote apoptosis, underscoring its potential application in cancer therapy. Further preclinical and clinical studies are warranted to validate these results and explore the full therapeutic potential of HG110 in STAT3-driven malignancies.

Direct binding to the active site of STAT3 can interfere with its phosphorylation at Tyr705, subsequently affecting cellular activities. In cancer cells, receptor-bound JAK kinases, primarily JAK1 and JAK2, are activated upon cytokine signaling and phosphorylate the cytoplasmic domain of the receptor [47]. This phosphorylation generates docking sites for STAT3, allowing it to bind via its SH2 domain. The receptor-associated JAKs then phosphorylate STAT3 at Tyr705, which induces its dimerization and translocation to the nucleus [48]. Once in the nucleus, STAT3 modulates the expression of various target genes involved in inflammatory responses, cancer progression, and immune evasion [47]. Several studies have demonstrated the potential of targeting STAT3 phosphorylation as a therapeutic strategy: Zhang et al. reported [49] that S3I-201.1066 binds to the STAT3 SH2 domain and disrupts STAT3 phosphorylation in vivo; Gabriele et al. [50] found that thiosulfonic compounds selectively bind the STAT3 SH2 domain, effectively inhibiting cancer cell proliferation; and He et al. [51] described the use of oral STAT3 inhibitors to suppress gastric cancer proliferation by downregulating STAT3 phosphorylation at Tyr705. These findings highlight the growing promise of targeting the STAT3 active site with emerging inhibitors as a therapeutic strategy for cancer.

Traditional drug discovery methodologies have often been hindered by high costs, low efficiency, and a dependence on the experience and intuition of medicinal chemists, which can limit the exploration of chemical space. Consequently, scaffold designs and structural modifications are typically constrained to well-known chemical entities. In contrast, the advent of machine learning (ML)-aided drug design has revolutionized the field by enabling a more extensive exploration of chemical space and enhancing the hit rate of targeted compounds. By leveraging sophisticated algorithms and comprehensive chemical datasets, ML models can predict the biological activities and physicochemical properties of molecules with high accuracy, significantly expediting the drug discovery process.

Various machine learning architectures, including Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs), have demonstrated remarkable potential in drug design. These models can efficiently handle complex chemical data, predict molecular properties, and forecast interactions with biological targets. In particular, generative models have introduced novel approaches to drug discovery by learning the probability distribution of training data, extracting key structural features, and generating novel compounds with the desired properties. This capability allows for the generation of new molecules that resemble known active compounds but possess unique structural variations that could lead to improved efficacy and reduced toxicity. In our investigation, we introduced the GDL to build the potential STAT3 inhibitors, which were trained from ZINC database and fine-tuned by current STAT3 inhibitors (Fig. 3A). This virtual molecular library contained the enough chemical space exploration and provided more insights to develop emerging chemical entities as potential STAT3 inhibitors. After screening, we selected the potential molecules, which contained the benzotriazole scaffold and modification with various chemical space exploration.

The application of machine learning extends beyond molecule generation and encompasses a wide range of drug discovery activities, including predicting drug–protein interactions, assessing drug efficacy, and identifying safety biomarkers. These capabilities streamline complex workflows and increase the throughput of drug discovery pipelines. However, despite the significant advancements, several challenges persist. The success of ML models is contingent on the quality and diversity of training data, and there is an ongoing need to enhance the interpretability and transparency of these models to facilitate their integration into conventional drug discovery workflows.

The newly identified STAT3 inhibitors, including HG110, offer a promising therapeutic strategy for NSCLC by targeting the pivotal STAT3 signaling pathway, which plays a critical role in tumorigenesis, immune evasion, and metastasis [47]. These inhibitors work by directly binding to the SH2 domain or the active site of STAT3, preventing its phosphorylation at Tyr705, a crucial step for STAT3 activation and its subsequent nuclear translocation [50]. This inhibition disrupts STAT3's ability to regulate gene expression involved in cell proliferation, survival, and immune modulation, offering a novel therapeutic approach. In comparison to current standard therapies, such as chemotherapy, EGFR inhibitors, and immune checkpoint inhibitors, STAT3 inhibitors provide several potential advantages. Chemotherapy, while effective, is often associated with significant off-target toxicity and limited efficacy in advanced or metastatic NSCLC [52]. EGFR inhibitors and ALK inhibitors target specific mutations but are limited by the development of resistance [53]. Immune checkpoint inhibitors, though revolutionary in treating NSCLC, are not effective in all patients and may result in immune-related adverse events [54]. STAT3 inhibitors, by targeting a central oncogenic pathway implicated in both tumor progression and resistance mechanisms, may overcome some of these limitations. Moreover, preliminary studies suggest these inhibitors are well tolerated with a manageable safety profile, presenting fewer systemic side effects compared to chemotherapy and fewer immune-related adverse events compared to checkpoint inhibitors. However, challenges remain, including the potential for resistance to STAT3 inhibition and the need for further clinical validation. When used in combination with other therapies, such as immune checkpoint inhibitors, STAT3 inhibitors could potentially enhance therapeutic outcomes, particularly in patients with resistant or refractory disease. Overall, while current therapies have provided significant survival benefits, the emerging STAT3 inhibitors represent a promising addition to the therapeutic armamentarium for NSCLC, with the potential to improve patient outcomes, especially in cases of resistance to traditional treatment modalities.

Conclusion

This study demonstrates that HG110, a novel small-molecule inhibitor, effectively suppresses NSCLC cell proliferation and induces apoptosis in vitro by targeting STAT3 phosphorylation. Mechanistically, HG110 inhibits STAT3 phosphorylation at Tyr705, disrupting STAT3-mediated oncogenic signaling pathways critical to tumor progression. In vivo, HG110 significantly reduces tumor growth in NSCLC xenograft models, underscoring its potential as an antitumor agent. These findings highlight HG110 as a promising therapeutic candidate for NSCLC treatment. Future research should prioritize optimizing its pharmacokinetic and pharmacodynamic properties to improve clinical applicability. Additionally, investigating the synergistic effects of HG110 in combination with current standard therapies—such as targeted therapies and immune checkpoint inhibitors—may further elucidate its broader therapeutic potential. In conclusion, this study provides a robust foundation for the continued development of HG110 as a novel and effective STAT3 inhibitor, addressing an important unmet need in NSCLC management.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11030-024-11067-5.

Acknowledgements We are grateful to Professor Chen Yihua's research group for providing us with the small-molecule compounds. This work was supported by the National Natural Science Foundation of China (82060663, 82260716), Regional Key R&D Program of Ningxia Hui Automous Region(2023BEG02010).

Author contributions Weiji Cai, Beier Jiang, Yichen Ying, Lei Ma, Jing Chen, and Tao Li designed and performed in vivo experiments; Beier Jiang contributed to machine learning- and molecular dynamics simulation-associated experiments; Weiji Cai, Beier Jiang, Yichen Ying, and Lei Ma provided cell lines and assisted with in vitro experiments; Beier Jiang and Weiji Cai analyzed the data and discussed the results; Jing Chen and Tao Li conceived and co-supervised the overall project and wrote the paper with contributions from all authors.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Alexander M, Kim SY, Cheng H (2020) Update 2020: management of non-small cell lung cancer. Lung 198:897–907. https://doi.org/10.1007/s00408-020-00407-5
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA (2008) Non-small cell lung cancer: epidemiology, risk factors, treatment, and

survivorship. Paper Present Mayo Clin Proc 83:584–594. https://doi.org/10.4065/83.5.584
3. Rath B, Plangger A, Hamilton G (2020) Non-small cell lung cancer-small cell lung cancer transformation as mechanism of resistance to tyrosine kinase inhibitors in lung cancer. Cancer Drug Resist 3(171–1783):171. https://doi.org/10.20517/cdr.2019.85
4. Siegel RL, Giaquinto AN, Jemal A (2024) Cancer statistics, 2024. CA Cancer J Clin 74:12–49. https://doi.org/10.3322/caac.21820
5. Singh R, Manna S, Nandanwar H, Purohit R (2024) Bioactives from medicinal herb against bedaquiline resistant tuberculosis: removing the dark clouds from the horizon. Microbes Infect 26:105279. https://doi.org/10.1016/j.micinf.2023.105279
6. Tanwar G, Purohit R (2019) Gain of native conformation of Aurora A S155R mutant by small molecules. J Cell Biochem 120:11104–11114. https://doi.org/10.1002/jcb.28387
7. Bhardwaj V, Singh R, Singh P, Purohit R, Kumar S (2020) Elimination of bitter-off taste of stevioside through structure modification and computational interventions. J Theor Biol 486:110094. https://doi.org/10.1016/j.jtbi.2019.110094
8. Rajasekaran R, George Priya Doss C, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R (2008) Effect of deleterious nsSNP on the HER2 receptor based on stability and binding affinity with herceptin: a computational approach. CR Biol 331:409–417. https://doi.org/10.1016/j.crvi.2008.03.004
9. Imyanitov EN, Iyevleva AG, Levchenko EV (2021) Molecular testing and targeted therapy for non-small cell lung cancer: current status and perspectives. Crit Rev Oncol Hematol 157:103194. https://doi.org/10.1016/j.critrevonc.2020.103194
10. Passaro A, Leighl N, Blackhall F, Popat S, Kerr K, Ahn M, Arcila M, Arrieta O, Planchard D, De Marinis F (2022) ESMO expert consensus statements on the management of EGFR mutant non-small-cell lung cancer. Ann Oncol 33:466–487. https://doi.org/10.1016/j.annonc.2022.02.003
11. Drilon A, Camidge DR, Lin JJ, Kim S-W, Solomon BJ, Dziadziuszko R, Besse B, Goto K, de Langen AJ, Wolf J (2024) Repotrectinib in ROS1 fusion–positive non–small-cell lung cancer. N Engl J Med 390:118–131. https://doi.org/10.1056/nejmoa2302299
12. Halani V, Sharayah A, Beck B, Patolia S (2024) "New targets in non–small-cell lung cancer"—RET, HER2, and KRAS. Am J Respir Crit Care Med 209:748–750. https://doi.org/10.1164/rccm.202208-1596rr
13. Uehara Y, Watanabe K, Hakozaki T, Yomota M, Hosomi Y (2022) Efficacy of first-line immune checkpoint inhibitors in patients with advanced NSCLC with KRAS, MET, FGFR, RET, BRAF, and HER2 alterations. Thorac Cancer 13:1703–1711. https://doi.org/10.1111/1759-7714.14448
14. Huang M-Y, Jiang X-M, Wang B-L, Sun Y, Lu J-J (2021) Combination therapy with PD-1/PD-L1 blockade in non-small cell lung cancer: strategies and mechanisms. Pharmacol Ther 219:107694. https://doi.org/10.1016/j.pharmthera.2020.107694
15. Liu Z, Ma L, Sun Y, Yu W, Wang X (2021) Targeting STAT3 signaling overcomes gefitinib resistance in non-small cell lung cancer. Cell Death Dis 12:561. https://doi.org/10.1038/s41419-021-03844-z
16. Yu H, Lee H, Herrmann A, Buettner R, Jove R (2014) Revisiting STAT3 signalling in cancer: new and unexpected biological functions. Nat Rev Cancer 14:736–746. https://doi.org/10.1038/nrc3818
17. Liu Z-y, Zhang Y-w, Zhuang H-x, Ou Y-j, Jiang Q-y, Li P-f, He Y-m, Ren Y, Mao X-l (2024) Inhibiting the otub1/phosphorylated STAT3 axis for the treatment of non-small cell lung cancer. Acta Pharmacol Sin. https://doi.org/10.1038/s41401-024-01366-w
18. Dong J, Cheng X-D, Zhang W-D, Qin J-J (2021) Recent update on development of small-molecule STAT3 inhibitors for cancer therapy: from phosphorylation inhibition to protein degradation. J Med Chem 64:8884–8915. https://doi.org/10.1021/acs.jmedchem.1c00629
19. Wang Z, Hui C, Xie Y (2021) Natural STAT3 inhibitors: a mini perspective. Bioorg Chem 115:105169. https://doi.org/10.1016/j.bioorg.2021.105169
20. Zhang C-C, Wu T, Guan L, Wang Y-J, Yao R-Q, Gao D-S, Li F (2022) Effects of STAT3 inhibitor BP-1-102 on the proliferation, invasiveness, apoptosis and neurosphere formation of glioma cells in vitro. Cell Biochem Biophys 80:723–735. https://doi.org/10.1007/s12013-022-01088-y
21. Brambilla L, Lahiri T, Cammer M, Levy DE (2020) STAT3 inhibitor OPB-51602 is cytotoxic to tumor cells through inhibition of complex I and ROS induction. iScience 23:101822. https://doi.org/10.1016/j.isci.2020.101822
22. Liang D, Wang Q, Zhang W, Tang H, Song C, Yan Z, Liang Y, Wang H (2024) JAK/STAT in leukemia: a clinical update. Mol Cancer 23:25. https://doi.org/10.1186/s12943-023-01929-1
23. Du Y, Jamasb AR, Guo J, Fu T, Harris C, Wang Y, Duan C, Liò P, Schwaller P, Blundell TL (2024) Machine learning-aided generative molecular design. Nat Mach Intell. https://doi.org/10.1038/s42256-024-00843-5
24. Arooj M, Shehadi I, Nassab CN, Mohamed AA (2022) Computational insights into binding mechanism of drugs as potential inhibitors against SARS-CoV-2 targets. Chem Pap 76:111–121. https://doi.org/10.1007/s11696-021-01843-0
25. Arooj M, Kim S, Sakkiah S, Cao GP, Lee Y, Lee KW (2013) Molecular modeling study for inhibition mechanism of human chymase and its application in inhibitor design. PLoS ONE 8:e62740. https://doi.org/10.1371/journal.pone.0062740
26. Martinelli DD (2022) Generative machine learning for de novo drug discovery: a systematic review. Comput Biol Med 145:105403. https://doi.org/10.1016/j.compbiomed.2022.105403
27. Gou R, Yang J, Guo M, Chen Y, Xue W (2024) CNSMolGen: a bidirectional recurrent neural network-based generative model for de novo central nervous system drug design. J Chem Inf Model 64:4059–4070. https://doi.org/10.1021/acs.jcim.4c00504
28. Weller JA, Rohs R (2024) Structure-based drug design with a deep hierarchical generative model. J Chem Inf Model 64:6450–6463. https://doi.org/10.1021/acs.jcim.4c01193
29. Lv Q, Zhou F, Liu X, Zhi L (2023) Artificial intelligence in small molecule drug discovery from 2018 to 2023: does it really work? Bioorg Chem 141:106894. https://doi.org/10.1016/j.bioorg.2023.106894
30. Jiang R, Gogineni T, Kammeraad J, He Y, Tewari A, Zimmerman PM (2022) Conformer-RL: a deep reinforcement learning library for conformer generation. J Comput Chem 43:1880–1886. https://doi.org/10.1002/jcc.26984
31. Arooj M, Sakkiah S, Cao GP, Kim S, Arulalapperumal V, Lee KW (2015) Finding off-targets, biological pathways, and target diseases for chymase inhibitors via structure-based systems biology approach. Proteins 83:1209–1224. https://doi.org/10.1002/prot.24677
32. Liao J, Wu M, Gao J, Chen C (2024) Calculation of solvation force in molecular dynamics simulation by deep-learning method. Biophys J 123:2830–2838. https://doi.org/10.1016/j.bpj.2024.02.029
33. Tam B, Qin Z, Zhao B, Wang SM, Lei CL (2023) Integration of deep learning with Ramachandran plot molecular dynamics simulation for genetic variant classification. iScience 26:106122. https://doi.org/10.1016/j.isci.2023.106122
34. Zhou G, Rusnac D-V, Park H, Canzani D, Nguyen HM, Stewart L, Bush MF, Nguyen PT, Wulff H, Yarov-Yarovoy V (2024) An artificial intelligence accelerated virtual screening platform for drug discovery. Nat Commun 15:7761. https://doi.org/10.1038/s41467-024-52061-7

35. Li Y, Zhang L, Wang Y, Zou J, Yang R, Luo X, Wu C, Yang W, Tian C, Xu H (2022) Generative deep learning enables the discovery of a potent and selective RIPK1 inhibitor. Nat Commun 13:6891. https://doi.org/10.1038/s41467-022-34692-w
36. Bai L, Zhou H, Xu R, Zhao Y, Chinnaswamy K, McEachern D, Chen J, Yang CY, Liu Z, Wang M, Liu L, Jiang H, Wen B, Kumar P, Meagher JL, Sun D, Stuckey JA, Wang S (2019) A potent and selective small-molecule degrader of STAT3 achieves complete tumor regression in vivo. Cancer Cell 36:498-511.e417. https://doi.org/10.1016/j.ccell.2019.10.002
37. Gupta A, Purohit R (2024) Identification of potent BRD4-BD1 inhibitors using classical and steered molecular dynamics based free energy analysis. J Cell Biochem 125:e30532. https://doi.org/10.1002/jcb.30532
38. Cao Y, Li L (2014) Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model. Bioinformatics 30:1674–1680. https://doi.org/10.1093/bioinformatics/btu104
39. Yang X, Liu Y, Gan J, Xiao ZX, Cao Y (2022) FitDock: protein-ligand docking by template fitting. Brief Bioinform 23:bbac087. https://doi.org/10.1093/bib/bbac087
40. McInnes L, Healy J,Melville J (2018) Umap: uniform manifold approximation and projection for dimension reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426
41. Rugard M, Jaylet T, Taboureau O, Tromelin A, Audouze K (2021) Smell compounds classification using UMAP to increase knowledge of odors and molecular structures linkages. PLoS ONE 16:e0252486. https://doi.org/10.1371/journal.pone.0252486
42. Yang T, Huo J, Xu R, Su Q, Tang W, Zhang D, Zhu M, Zhan Y, Dai B, Zhang Y (2021) Selenium sulfide disrupts the PLAGL2/C-MET/STAT3-induced resistance against mitochondrial apoptosis in hepatocellular carcinoma. Clin Transl Med 11:e536. https://doi.org/10.1002/ctm2.536
43. Jia Y, Wang Q, Liang M, Huang K (2022) KPNA2 promotes angiogenesis by regulating STAT3 phosphorylation. J Transl Med 20:627. https://doi.org/10.1186/s12967-022-03841-6
44. Liu Y, Xu Q, Deng F, Zheng Z, Luo J, Wang P, Zhou J, Lu X, Zhang L, Chen Z, Zhang Q, Chen Q, Zuo D (2023) HERC2 promotes inflammation-driven cancer stemness and immune evasion in hepatocellular carcinoma by activating STAT3 pathway. J Exp Clin Cancer Res : CR 42:38. https://doi.org/10.1186/s13046-023-02609-0
45. Yu H, Pardoll D, Jove R (2009) STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer 9:798–809. https://doi.org/10.1038/nrc2734
46. Zegeye MM, Lindkvist M, Fälker K, Kumawat AK, Paramel G, Grenegård M, Sirsjö A, Ljungberg LU (2018) Activation of the JAK/STAT3 and PI3K/AKT pathways are crucial for IL-6 trans-signaling-mediated pro-inflammatory response in human vascular endothelial cells. Cell Commun Signal 16:55. https://doi.org/10.1186/s12964-018-0268-4
47. Johnson DE, O'Keefe RA, Grandis JR (2018) Targeting the IL-6/JAK/STAT3 signalling axis in cancer. Nat Rev Clin Oncol 15:234–248. https://doi.org/10.1038/nrclinonc.2018.8
48. Genini D, Brambilla L, Laurini E, Merulla J, Civenni G, Pandit S, D'Antuono R, Perez L, Levy DE, Pricl S, Carbone GM, Catapano CV (2017) Mitochondrial dysfunction induced by a SH2 domain-targeting STAT3 inhibitor leads to metabolic synthetic lethality in cancer cells. Proc Natl Acad Sci USA 114:E4924-e4933. https://doi.org/10.1073/pnas.1615730114
49. Zhang X, Yue P, Fletcher S, Zhao W, Gunning PT, Turkson J (2010) A novel small-molecule disrupts Stat3 SH2 domain-phosphotyrosine interactions and Stat3-dependent tumor processes. Biochem Pharmacol 79:1398–1409. https://doi.org/10.1016/j.bcp.2010.01.001
50. Gabriele E, Ricci C, Meneghetti F, Ferri N, Asai A, Sparatore A (2017) Methanethiosulfonate derivatives as ligands of the STAT3-SH2 domain. J Enzyme Inhib Med Chem 32:337–344. https://doi.org/10.1080/14756366.2016.1252757
51. Chen H, Bian A, Zhou W, Miao Y, Ye J, Li J, He P, Zhang Q, Sun Y, Sun Z, Ti C, Chen Y, Yi Z, Liu M (2024) Discovery of the highly selective and potent STAT3 inhibitor for pancreatic cancer treatment. ACS Cent Sci 10:579–594. https://doi.org/10.1021/acscentsci.3c01440
52. Zhou C, Tang KJ, Cho BC, Liu B, Paz-Ares L, Cheng S, Kitazono S, Thiagarajan M, Goldman JW, Sabari JK, Sanborn RE, Mansfield AS, Hung JY, Boyer M, Popat S, Mourão Dias J, Felip E, Majem M, Gumus M, Kim SW, Ono A, Xie J, Bhattacharya A, Agrawal T, Shreeve SM, Knoblauch RE, Park K, Girard N (2023) Amivantamab plus chemotherapy in NSCLC with EGFR exon 20 insertions. N Engl J Med 389:2039–2051. https://doi.org/10.1056/NEJMoa2306441
53. Cooper AJ, Sequist LV, Lin JJ (2022) Third-generation EGFR and ALK inhibitors: mechanisms of resistance and management. Nat Rev Clin Oncol 19:499–514. https://doi.org/10.1038/s41571-022-00639-9
54. Passaro A, Brahmer J, Antonia S, Mok T, Peters S (2022) Managing resistance to immune checkpoint inhibitors in lung cancer: treatment and novel strategies. J Clin Oncol 40:598–610. https://doi.org/10.1200/jco.21.01845