

MolGPT-assisted exploration of VEGFR2-related chemical space through virtual screening, experimental bridging, and molecular dynamics simulation

Taiying Li

Shandong University of Traditional Chinese Medicine

Shuning Diao

Shandong University of Traditional Chinese Medicine

Yuxiang He

Shandong University of Traditional Chinese Medicine

Shengzhen Hou

Shandong University of Traditional Chinese Medicine

Jinping Zhang

zhang.jinpingdoctor@126.com

Affiliated Hospital of Shandong University of Traditional Chinese Medicine

Article

Keywords: MolGPT, VEGFR2, generative molecular design, virtual screening, molecular dynamics simulations, kinase assay

Posted Date: October 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11092477/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

MolGPT-assisted exploration of VEGFR2-related chemical space through virtual screening, experimental bridging, and molecular dynamics simulation

Taiying Li¹, Shuning Diao¹, Yuxiang He¹, Shengzhen Hou¹,
Jinping Zhang^{2*}

¹First Clinical Medical College, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China.

²Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China.

*Corresponding author(s). E-mail(s): zhangjinpingdoctor@126.com;

Abstract

Vascular endothelial growth factor receptor 2 (VEGFR2) is a central regulator of tumor angiogenesis and an established therapeutic target in anticancer drug discovery. Generative molecular models provide a promising means of exploring chemical space beyond the limits of conventional compound libraries. However, translating de novo-generated candidates into experimentally testable compounds remains challenging because prioritized molecules may be unavailable or difficult to synthesize. Herein, we developed a candidate prioritization workflow combining MolGPT-based generation, graph convolutional network (GCN) activity prediction, molecular docking, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) assessment. The workflow further incorporated similarity matching to experimentally accessible surrogates, in vitro VEGFR2 kinase assays, molecular dynamics (MD) simulations, and MM-PBSA analysis. MolGPT generated 5,000 molecules, including 4,992 unique valid structures. GCN screening retained 522 candidates, docking prioritized nine molecules, and ADMET and drug-likeness assessment selected four candidates for similarity matching. Six commercially available surrogate compounds were subsequently tested in vitro. Levomefolic acid showed a non-monotonic concentration-response profile, with mean inhibition of 56.6% at 20 μ M decreasing to 12.8% at 40 μ M. This profile did not support a reliable quantitative potency estimate. Subsequent docking, MD simulations, and MM-PBSA analyses generated structural

hypotheses but did not establish VEGFR2 binding or validate the activity of the corresponding generated molecule. This study presents a practical strategy for connecting generative molecular design with exploratory experimental assessment of accessible compounds.

Keywords: MolGPT, VEGFR2, generative molecular design, virtual screening, molecular dynamics simulations, kinase assay

1 Introduction

According to the International Agency for Research on Cancer, nearly 20 million new cancer cases and approximately 9.7 million cancer deaths occurred worldwide in 2022. New cases are projected to reach 35 million annually by 2050, an increase of 77% from 2022 [6]. Tumor angiogenesis drives tumor growth and metastasis, and induction of angiogenesis is considered a hallmark of cancer [13]. Once a tumor exceeds the diffusion limit of several millimeters, it must activate the “angiogenic switch” to form a new vascular network [7]. Consequently, antiangiogenic therapy, which targets and suppresses this change, has become an essential therapeutic strategy in modern oncology [19]. Vascular endothelial growth factor receptor 2 (VEGFR2) is the principal transmembrane receptor that mediates proangiogenic signaling and endothelial-cell proliferation and migration. VEGFR1 mainly regulates macrophage migration, while VEGFR3 is predominantly involved in lymphangiogenesis. Selective inhibition of VEGFR2 kinase activity is therefore a widely recognized strategy for suppressing tumor angiogenesis [22].

Clinically used VEGFR2 inhibitors, including sorafenib and sunitinib, commonly contain aryl urea, quinoline or indole scaffolds and inhibit VEGFR2 by occupying their adenosine triphosphate (ATP) binding pocket. However, they also inhibit kinases such as platelet-derived growth factor receptors and KIT, contributing to toxicities, including hypertension and hand-foot skin reaction, and restricting their therapeutic windows [24, 30]. Prolonged treatment may also promote tumor escape through alternative fibroblast growth factor and platelet-derived growth factor pathways. This “angiogenic redundancy” can contribute to acquired resistance [10]. These limitations motivate the discovery of candidate VEGFR2 inhibitors with distinct scaffolds and binding characteristics, together with early assessment of drug-like properties, target relevance, and potential safety.

Traditional high-throughput screening has been successful but remains costly and time-consuming, and its accessible chemical space is limited by the size and diversity of physical compound libraries. Deep-learning-based artificial intelligence, particularly generative models, offers an alternative approach to these constraints [31]. Unlike virtual screening of existing libraries, generative models can design molecular structures de novo. Encoding molecules as simplified molecular input line entry system (SMILES) strings allow natural language processing architectures to be adapted for molecular generation. Recurrent neural networks and adversarial generative networks have demonstrated this potential [5]. For example, long- and short-term memory

and adversarial generative network models generated novel candidate inhibitors predicted to bind to the main protease SARS-CoV-2 [3, 8]. Transformer architectures can capture long-range sequence dependencies [34]. MolGPT applies a Transformer decoder, next-token prediction, and masked self-attention to model molecular syntax and higher-order structural patterns. Hence, it may generate novel SMILES strings that encode chemically plausible structures [4].

Computational methods have also advanced the discovery of VEGFR2 inhibitors. Hou et al. applied a graph neural network to screen traditional Chinese medicine monomers, and subsequent *in vitro* kinase assays identified three natural products with inhibitory activity of VEGFR2 [14]. Liu et al. used linked fragments for *de novo* design, but molecular diversity and novelty were constrained by the predefined fragment library. Their linear workflow also lacked deep-learning-based property prediction [18]. Nayariseri et al. generated tens of millions of molecules targeting the broader VEGFR family, and docking and molecular dynamics (MD) simulations suggested the potential of selected candidates. However, this approach did not provide extensive VEGFR2-specific optimization or selective design [25]. Thus, insufficient selectivity, adverse effects, acquired resistance, and limited scaffold innovation remain challenges in the discovery of VEGFR2 inhibitors.

In this study, we developed a multistage candidate-prioritization workflow integrating molecular generation using the MolGPT model, VEGFR2 activity prediction based on graph convolutional networks (GCN), virtual screening based on structures and drug-like property assessment. Molecular Sets (MOSES) metrics and chemical-space analyzes evaluated the validity, uniqueness, novelty, and structural diversity of the MolGPT-generated molecular library. The previously developed and evaluated GCN assigned each molecule a GCN prediction score, which was used for first-stage computational prioritization. The selected candidates then underwent molecular docking, absorption, distribution, metabolism, excretion, and toxicity prediction (ADMET) and drug-like property assessment. As some prioritized MolGPT-generated molecules were unavailable for direct purchase or rapid synthesis, an accessible compound library was searched for similarity-matched surrogate compounds. These surrogates, rather than the original MolGPT-generated molecules, underwent exploratory *in vitro* VEGFR2 kinase tests. Compounds showing preliminary experimental responses were further examined by molecular docking, MD simulations, and binding free-energy analysis. These analyzes explored potential binding modes and computational indicators of complex behavior. The general workflow is shown in Figure 1.

2 Materials and methods

2.1 MolGPT

The MolGPT model uses a Transformer decoder to represent molecular SMILES strings as discrete token sequences. Masked self-attention captures contextual dependencies among previously generated tokens [4, 32]. During autoregressive generation, MolGPT predicts the next token from the preceding SMILES tokens. The attention operation is expressed as

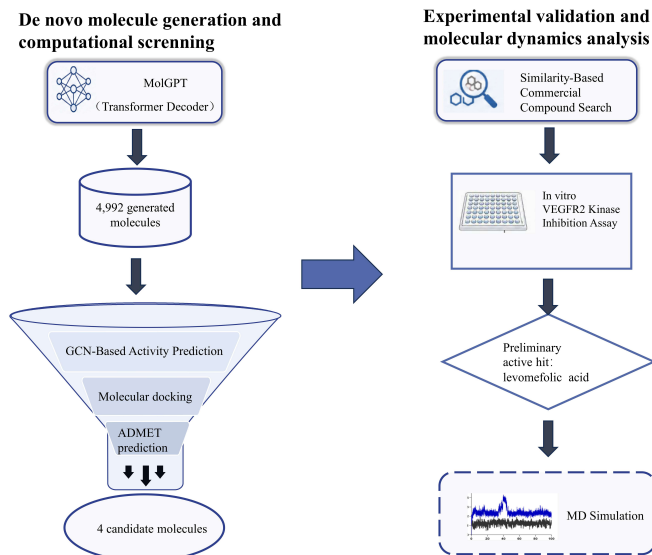


Fig. 1 Overall workflow of the MolGPT-assisted VEGFR2 candidate inhibitor prioritization strategy.

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} + M \right) V, \quad (1)$$

where Q , K , and V denote the query, key, and value matrices, respectively. The parameter d_k is the dimensionality of the key vectors, while M is the attention mask that prevents access to subsequent tokens. Given the hidden state of the decoder h_t , the probability distribution of the next-token is calculated as

$$P(x_t | x_{<t}) = \text{softmax}(Wh_t + b). \quad (2)$$

Thus, MolGPT performs autoregressive sequence modeling conditioned on previously generated SMILES fragments. In this study, the MolGPT model served as a general-purpose de novo molecular generator and was not specifically fine-tuned for VEGFR2. It was used to construct a virtual molecular library intended to encompass structural novelty and chemical diversity. The MolGPT-generated molecular library provided the candidate chemical space for downstream prediction and multistage screening of VEGFR2 activity.

2.2 Chemical space analysis

To compare the chemical-space distributions of the MolGPT-generated molecules and the dataset used for development of the VEGFR2 GCN model, we performed a comparative chemical-space analysis [4]. The model-development dataset comprised 5,564 molecules, including 2,841 active and 2,723 inactive compounds [14]. RDKit was used to calculate a 2,048-bit extended-connectivity fingerprint (ECFP4), corresponding to a Morgan fingerprint with a radius of 2, for each molecule [28].

Uniform Manifold Approximation and Projection (UMAP) was applied to the fingerprint vectors for nonlinear dimensionality reduction and two-dimensional visualization [21]. The Python `umap-learn` package was used with `n_neighbors` = 15 and `min_dist` = 0.1; all other parameters were retained at their default values. The MolGPT-generated molecules and the active and inactive compounds from the model-development dataset were projected into the same two-dimensional embedding. Their distributions were visualized to compare the relative positions of the generated molecules within the chemical space represented by the original GCN dataset. In addition to UMAP visualization, structural similarity to the model-development dataset was quantified directly using the ECFP4 fingerprints. For each generated compound, the maximum Tanimoto similarity across the reference compounds was recorded as T_{max} . This analysis was performed separately for the 4,992 unique generated compounds and the 522 GCN-selected compounds. For descriptive reporting, compounds were grouped into low-similarity ($T_{\text{max}} < 0.5$), intermediate-similarity ($0.5 \leq T_{\text{max}} < 0.6$), and high-similarity ($T_{\text{max}} \geq 0.6$) categories. These thresholds were descriptive cutoffs, not validated boundaries of the GCN model’s applicability domain. Similarity distributions were summarized using the mean, standard deviation (SD; calculated with a denominator of N), median, and range.

2.3 Assessment of molecular quality and diversity using the MOSES benchmark

To evaluate the molecular generation performance of the MolGPT model, we used the Molecular Sets (MOSES) benchmarking platform [27]. MOSES is based on the ZINC Clean Leads dataset and provides standardized training and test sets with metrics for molecular validity, novelty, diversity, and distributional similarity. MolGPT was compared with the MOSES baseline models, including a character-level recurrent neural network (CharRNN), a variational autoencoder (VAE), an adversarial autoencoder (AAE), and a junction-tree variational autoencoder (JTN-VAE).

Validity, Uniqueness, and Novelty assessed the generation of valid, nonredundant, and novel molecular structures, respectively. Internal Diversity (IntDiv) quantified structural diversity within the MolGPT-generated molecular library. Scaffold Similarity (Scaf) and Fragment Similarity (Frag) compared scaffold and fragment distributions between the generated library and the training set.

Fréchet ChemNet Distance (FCD) quantified the statistical distance between molecules generated and set up for training in a ChemNet-derived feature space. Similarity to Nearest Neighbor (SNN) measured the similarity of each generated molecule to its nearest neighbor in the reference set. The filter metric quantified the proportion of molecules generated passing medicinal chemistry filters (MCF) and pan-assay interference compounds (PAINS) filters.

2.4 Virtual screening based on a graph neural network

To prioritize compounds with potential VEGFR2 inhibitory activity, we applied a graph convolutional network (GCN) previously developed and systematically evaluated by our group [14]. The model was constructed using VEGFR2-related activity

data obtained from the ChEMBL database. After retaining records containing SMILES and IC₅₀ data and removing duplicates, 5,564 compounds remained. Compounds with IC₅₀ values below 500 nM were classified as active, whereas those with IC₅₀ values greater than or equal to 500 nM were classified as inactive, resulting in 2,841 active and 2,723 inactive compounds. In the original study, the dataset was randomly divided into training, validation, and test sets in an approximate 8:1:1 ratio for model construction, hyperparameter optimization, and independent performance evaluation, respectively.

The previous study compared random forest, support vector machine, XGBoost, graph attention network, GCN, and message-passing neural network models. On the independent test set, the GCN achieved an accuracy of 0.8276, sensitivity of 0.8015, specificity of 0.8526, Matthews correlation coefficient of 0.6554, and receiver operating characteristic area under the curve of 0.8937. Based on its overall performance, the previously developed GCN was retained in the present study as a first-stage activity-prioritization model.

The GCN was not retrained in the present study. Its output scores were used for computational prioritization rather than as direct evidence of VEGFR2 inhibition. The MolGPT model generated 5,000 molecules. Following SMILES standardization, validity checking, and duplicate removal, 4,992 unique molecular structures were retained. These molecules were converted into molecular graphs using the graph-construction and feature-encoding procedures described in the previous study, and a GCN-predicted activity score was obtained for each molecule.

To characterize the structural coverage of the generated compounds relative to the original GCN dataset, a similarity-based chemical-space coverage analysis was performed. The curated dataset of 5,564 VEGFR2-related compounds used for construction and evaluation of the original GCN model served as the reference chemical space. Because the compound-level assignments of the original training, validation, and test subsets were unavailable, the complete model-development dataset was used as the structural reference set and was not treated as the actual GCN training subset.

Molecular structures were encoded as Morgan fingerprints with a radius of 2 and a length of 2,048 bits using RDKit. These fingerprints were considered ECFP4-like representations. For each generated compound, Tanimoto similarity was calculated against every compound in the reference dataset, and the maximum Tanimoto coefficient ($T_{c_{max}}$) was retained as a measure of structural proximity to the chemical space represented by the original GCN dataset.

For descriptive assessment of structural coverage, the generated compounds were categorized into three similarity regions: high similarity ($T_{c_{max}} \geq 0.6$), intermediate similarity ($0.5 \leq T_{c_{max}} < 0.6$), and low similarity ($T_{c_{max}} < 0.5$). These similarity regions were used solely for descriptive characterization of structural proximity and were not calibrated against GCN prediction performance. Therefore, they should not be interpreted as validated thresholds defining the predictive applicability domain of the GCN model.

Because the generated library was designed to explore chemical structures beyond known VEGFR2-associated compounds, a GCN output score threshold of 0.5 was used for first-stage computational prioritization. Compounds with a GCN output score

greater than 0.5 were advanced to molecular docking. Given the structural novelty of the generated chemical space, particularly among compounds showing low similarity to the reference dataset, GCN output scores were interpreted as prioritization scores rather than as calibrated probabilities or definitive predictions of VEGFR2 inhibitory activity.

2.5 Molecular docking

Molecular docking was performed using AutoDock Vina 1.2.0 [9]. The VEGFR2 co-crystal structure with Protein Data Bank (PDB) ID 4ASE was obtained from the RCSB Protein Data Bank (<https://www.rcsb.org/>). The receptor was prepared using AutoDockTools [23]. Water molecules, the co-crystallized ligand, and other non-protein heteroatoms were removed. Polar hydrogen atoms were added, Gasteiger partial charges were assigned, and the prepared receptor was saved in PDBQT format.

Small-molecule ligands were batch-converted from SMILES strings into three-dimensional structures. The ligand structures were energy-minimized in Chem3D using the Merck molecular force field 94 (MMFF94). Gasteiger partial charges were subsequently assigned, and the processed ligands were saved in PDBQT format.

The docking region encompassed the ATP-binding pocket occupied by the co-crystallized ligand. A grid box measuring $25.612 \times 25.612 \times 25.612$ Å was placed in the geometric center of the original ligand. The center coordinates were $X = -22.922$, $Y = 0.075$, and $Z = -12.217$. The grid included the active pocket and adjacent regions. The receptor was treated as rigid, whereas the ligands were treated as flexible.

The exhaustiveness parameter was set to 8, with all other parameters retained at their default values. AutoDock Vina generated nine poses per ligand and ranked them using the Vina-predicted binding score in kcal/mol. The pose with the lowest predicted score was selected for subsequent interaction analysis and MD simulations. The docking results were analyzed and visualized using PyMOL 2.5 and related molecular visualization tools.

2.6 ADMET prediction

To further prioritize candidates according to their predicted developability, we conducted a third-stage screening using physicochemical criteria and ADMET. The screen included the nine compounds retained after GCN-based activity prioritization and molecular docking. Candidates were first evaluated according to Lipinski’s Rule of Five, and those not meeting the specified drug-likeness criteria were excluded [17].

The ADMET properties of the remaining molecules were predicted using ADMET-lab 3.0 (<https://admetlab3.scbdd.com/>) [11]. Molecules retained after these computational filters were advanced to similarity-based screening.

2.7 Similarity-based screening

The prioritized MolGPT-generated molecules were structurally novel, unavailable commercially, and considered challenging to synthesize directly. To bridge computational screening and exploratory experimental assessment, the TargetMol Natural

Product Library was searched for commercially available, structurally related compounds. These similarity-matched surrogate compounds, rather than the original MolGPT-generated molecules, were selected for exploratory in vitro activity assessment.

RDKit was used to calculate 2,048-bit ECFP4 fingerprints with a radius of 2 for the MolGPT-generated molecules and the compounds in the TargetMol Natural Product Library [28, 35]. Structural similarity was quantified using the Tanimoto coefficient T_c [29]:

$$T_c(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (3)$$

where A and B denote the fingerprint vectors of a candidate MolGPT-generated molecule and a natural product, respectively. Because de novo-generated structures may be underrepresented in existing medicinal chemistry collections, no stringent high-similarity cutoff was imposed. Guided by the concept of scaffold hopping [15], prioritization focused on natural products showing some correspondence in local structural features, key pharmacophore distributions, or overall molecular topology.

The similarity-matched surrogate compounds were not considered direct substitutes for the original MolGPT-generated molecules, and their experimental results cannot be directly extrapolated to the corresponding generated molecules. They were used only to explore whether the related chemical space might contain indications of VEGFR2 inhibitory activity warranting further investigation.

2.8 In vitro kinase inhibition assay

The inhibitory effects of six similarity-matched surrogate compounds on KDR (VEGFR2) kinase activity were evaluated using a LanthaScreen™ kinase assay. The final KDR and ATP concentrations were 0.05 nM and 100 μ M, respectively. KDR (VEGFR2) kinase (Carna, Cat. No. 08-191), Fluorescein-PolyGT substrate (Invitrogen, Cat. No. PV3610), and the LanthaScreen Tb-PY20 antibody kit (Invitrogen, Cat. No. PV3529) were used. Staurosporine was included as the positive control.

Stock solutions of the six test compounds were prepared at 40 mM in 100% DMSO and diluted to 4 mM working solutions. Ten-point, twofold serial dilutions were prepared in DMSO, yielding nominal final assay concentrations ranging from 40 to 0.078125 μ M. Staurosporine was tested using a separate ten-point, fourfold dilution series, with final concentrations ranging from 1,000 to approximately 0.003815 nM. A 200 nL aliquot of each dilution was transferred to a 384-well assay plate using an Echo 650 acoustic liquid handler. The DMSO concentration during the kinase reaction was approximately 1% (v/v).

The kinase buffer contained 50 mM HEPES (pH 7.5), 10 mM MgCl_2 , 0.01% Tween-20, 0.01% BSA, and 2 mM DTT. KDR kinase was diluted in the kinase buffer, and 10 μ L of the resulting kinase solution was added to each reaction well. No-enzyme control wells received an equal volume of kinase buffer without KDR. The plate was centrifuged at 1,000 rpm for 1 min and incubated at room temperature for 10 min.

Subsequently, 10 μ L of the ATP/substrate solution containing Fluorescein-PolyGT was added to each well. The plate was centrifuged again at 1,000 rpm for 1 min and

incubated at room temperature for 10 min. The reactions were terminated by adding 20 μL of $2\times$ stop solution containing the Tb-PY20 antibody. The plate was then centrifuged at 1,000 rpm for 1 min and incubated at room temperature for 60 min before signal acquisition.

Time-resolved fluorescence resonance energy transfer signals were measured using an EnVision 2104 Multilabel Reader. The 520/495-nm signal ratio was used to calculate percentage inhibition as follows:

$$\text{Inhibition (\%)} = \frac{\text{max} - \text{sample}}{\text{max} - \text{min}} \times 100, \quad (4)$$

where max represents the signal from the DMSO vehicle-control wells and min represents the signal from the no-enzyme control wells. Dose-response curves were fitted using XLFit Excel Add-in version 5.4.0.8 with a four-parameter logistic model:

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + (\text{IC}_{50}/X)^{\text{HillSlope}}}, \quad (5)$$

For levomefolic acid, the upper asymptote (Top) was fixed at 100% in the original laboratory fit. The fitted IC_{50} represents the midpoint between the fitted asymptotes and does not necessarily correspond to 50% absolute inhibition. The reported dataset comprised one plate-based assay run with two technical replicate wells per concentration. Inhibition values are reported as the mean $\pm$ sample standard deviation (SD) of the two wells.

2.9 Molecular dynamics simulations

All-atom molecular dynamics (MD) simulations were performed for 100 ns using GROMACS 2020.3-MODIFIED to characterize the dynamic behaviour and conformational stability of the protein-ligand complexes [1]. The protein was described using a CHARMM-family all-atom force field, whereas the ligands were represented using CHARMM-compatible parameters. Each protein-ligand complex was solvated in a cubic simulation box using the TIP3P water model. Na^+ and Cl^- ions were added to obtain electrically neutral systems with an ionic concentration of approximately 0.15 mol L^{-1} . The resulting systems contained approximately 68,000 atoms, including the protein, ligand, water molecules, and ions.

Prior to the production simulations, the systems were subjected to energy minimization followed by NVT and NPT equilibration.

Production simulations were performed using the md integrator with a time step of 2 fs for 50,000,000 integration steps under three-dimensional periodic boundary conditions. Nonbonded interactions were treated using the Verlet cutoff scheme. Long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method with a Fourier grid spacing of 0.16 nm, fourth-order interpolation, and an Ewald tolerance of 1×10^{-5} . The cutoff distances for short-range Coulomb and van der Waals interactions were both set to 1.0 nm. Long-range dispersion corrections were applied to both energy and pressure.

During the production simulations, the temperature was maintained at 300 K using the V-rescale thermostat with a coupling time constant of 0.1 ps. The pressure was

maintained at 1 bar using the Parrinello–Rahman barostat with isotropic pressure coupling, a coupling time constant of 2 ps, and a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. Bonds involving hydrogen atoms were constrained using the linear constraint solver (LINCS) algorithm with fourth-order expansion, whereas water geometry was constrained using the SETTLE algorithm. Trajectory coordinates and energies were recorded every 10 ps for subsequent analysis.

Root-mean-square deviation (RMSD), radius of gyration (R_g), solvent-accessible surface area (SASA), and protein–ligand hydrogen-bond profiles were calculated from the production trajectories. These descriptors were used to compare the dynamic and conformational behaviour of the two protein–ligand complexes.

Molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) calculations were subsequently performed using `g_mmpbsa` 2021.1 to estimate the relative binding energetics of the ligands with VEGFR2 [12, 16]. Snapshots were extracted at intervals of 100 ps from the 90–100 ns segment of each production trajectory, yielding 101 frames per complex. A single-trajectory protocol was applied. The molecular-mechanics electrostatic calculation used a solute dielectric constant of 2, and per-residue energy decomposition was performed. Polar and nonpolar solvation contributions were calculated using the Poisson–Boltzmann and SASA-based approaches implemented in `g_mmpbsa`, respectively. Entropic contributions were not included in the reported MM–PBSA energy estimates.

The resulting MM–PBSA values were interpreted as computational estimates of relative binding energetics rather than as direct measurements of experimental binding affinity or rigorous thermodynamic binding free energy. The binding-energy estimate was calculated as

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}). \quad (6)$$

3 Results

3.1 Chemical-space analysis of MolGPT-generated molecules

The Transformer decoder-based MolGPT model generated 5,000 small-molecule structures. A joint UMAP analysis compared these molecules with the chemical space represented by the 5,564 compounds in the original GCN model-development dataset. This dataset included 2,841 active and 2,723 inactive compounds.

As shown in Figure 2, most MolGPT-generated molecules were projected within the main region occupied by the original GCN dataset or its adjacent subregions. They overlapped substantially with both active and inactive compounds in the reference dataset, while some appeared in sparsely populated peripheral regions of the two-dimensional embedding. The active and inactive reference compounds were also not completely separated in the UMAP projection.

Within the limitations of this two-dimensional representation, the generated molecules did not appear to be broadly separated from the model-development dataset. However, apparent overlap in the UMAP projection did not necessarily indicate high pairwise similarity in the original fingerprint space. It also did not establish that the generated compounds were within a validated predictive applicability domain of the

GCN model. UMAP was therefore used only as a global visualization of the relative chemical-space distributions, whereas local structural proximity was evaluated separately using maximum Tanimoto similarity.

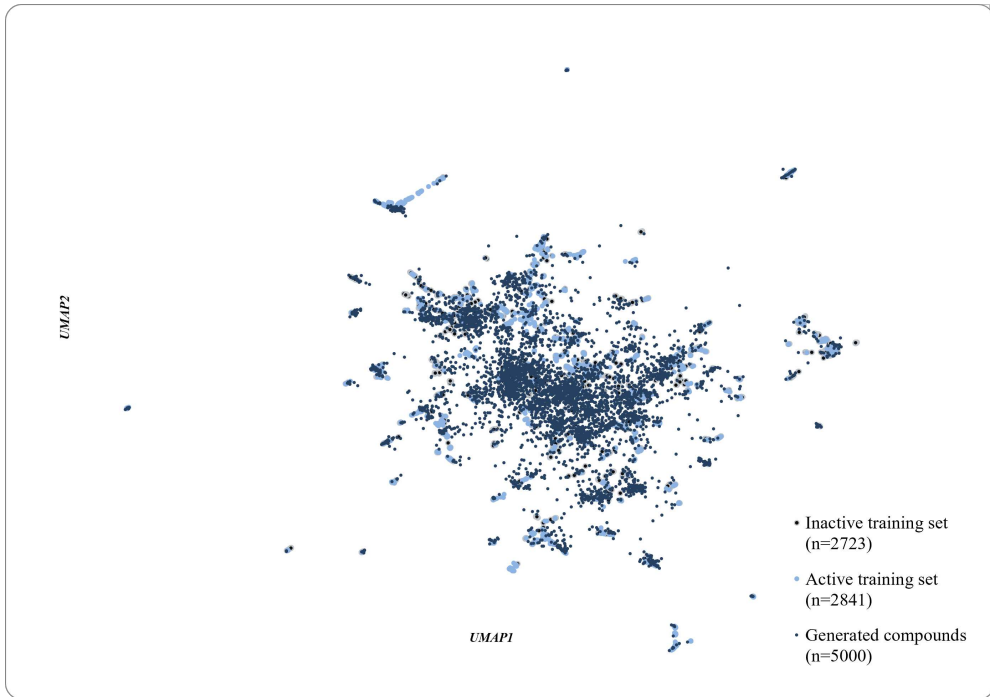


Fig. 2 Chemical-space distribution of MolGPT-generated molecules and the VEGFR2 GCN model-development dataset.

3.2 Assessment of molecular quality and diversity using the MOSES benchmark

The MOSES benchmark results are summarized in Table 1. MolGPT achieved a Validity score of 1.000 and a filter score of 0.993. All evaluated outputs met the MOSES validity criteria, and 99.3% passed the specified medicinal chemistry filters. These values were consistent with syntactically valid SMILES generation and a low frequency of structures flagged by the applied filters.

MolGPT achieved a Uniqueness score of 0.998 and a Novelty score of 0.824, comparable to the baseline models in Table 1. Under the MOSES definitions, these values indicated little duplication and a substantial proportion of structures absent from the reference training set. The SNN score was 0.609, within the range of classical models

such as CharRNN and AAE, indicating an intermediate level of nearest-neighbor similarity to the reference data set. However, SNN alone did not establish the applicability or reliability of subsequent activity predictions or virtual screening.

The 5,000-molecule sample used here was smaller than the larger generated sets commonly used in MOSES evaluations, potentially increasing the sensitivity of distribution-based metrics to sampling variation. Under this condition, MolGPT yielded an FCD of 1.986, higher than all baseline values in Table 1. This result indicated a greater statistical distance from the MOSES reference dataset in the ChemNet-derived feature space.

MolGPT also yielded a Fragment Similarity score of 0.145 and a Scaffold Similarity score of 0.756, the latter being lower than those of most baseline models. Together, these metrics indicated differences between the MolGPT-generated molecular library and the MOSES reference dataset in ChemNet-derived feature, fragment, and scaffold distributions. However, the lower Frag and Scaf values did not independently demonstrate greater molecular diversity or beneficial structural novelty.

The workflow was not intended to reproduce the unconditional distribution of the ZINC chemical space. Instead, the MolGPT model served as a general-purpose de novo molecular generator followed by property-oriented screening. Thus, deviations in distribution-matching metrics were not necessarily inconsistent with the objective of the study, but were not evidence of superior novelty. The lower Frag and Scaf scores may reflect both limited distributional matching and fragment or scaffold combinations that differ from the reference data set. Their potential relevance to structural differentiation or scaffold hopping requires direct scaffold-level analysis. These differences may also be associated with uncertainty in the synthetic accessibility and predicted ADMET properties, which requires a subsequent assessment of the candidate development potential.

Table 1 MOSES benchmark results for MolGPT and the baseline models.

Model	Validity (↑)	Uniqueness (↑)	Novelty (↑)	IntDiv (↑)	Scaf (↑)	FCD (↓)	Filters (↑)	SNN (↑)	Frag (↑)
MolGPT	1.000	0.998	0.824	0.852	0.756	1.986	0.993	0.609	0.145
CharRNN	0.975	0.999	0.842	0.856	0.924	0.073	0.994	0.601	1.000
VAE	0.977	0.998	0.695	0.856	0.939	0.099	0.997	0.626	0.999
AAE	0.937	0.997	0.793	0.856	0.902	0.556	0.996	0.608	0.991
JTN-VAE	1.000	0.999	0.914	0.855	0.896	0.395	0.976	0.547	0.997

3.3 Virtual screening results

Initial computational prioritization of the MolGPT-generated molecular library was performed using the GCN classification model for VEGFR2 inhibitors previously developed and evaluated by our group. Of the 5,000 generated molecules, eight duplicate structures were removed, leaving 4,992 unique compounds for GCN prediction. Compounds with a GCN output score greater than 0.5 were retained for subsequent

structure-based evaluation, yielding 522 candidates, corresponding to approximately 10.5% of the deduplicated molecular library.

A similarity-based chemical-space coverage analysis was subsequently performed for the 4,992 unique generated compounds using the 5,564 VEGFR2-related compounds in the original GCN model-development dataset as the reference chemical space. Overall, the generated molecules showed low structural similarity to the reference dataset. The mean maximum Tanimoto similarity, $T_{c_{\max}}$, was 0.298 ± 0.057 , with a median of 0.292 and a range of 0.141–0.667. Only 3 compounds (0.06%) were located in the high-similarity region ($T_{c_{\max}} \geq 0.6$), whereas 14 compounds (0.28%) were located in the intermediate-similarity region ($0.5 \leq T_{c_{\max}} < 0.6$). The remaining 4,975 compounds (99.66%) were located in the low-similarity region ($T_{c_{\max}} < 0.5$) relative to the reference chemical space.

The same analysis was then applied to the 522 compounds retained during GCN-based activity prioritization. Their mean $T_{c_{\max}}$ was 0.313 ± 0.059 , with a median of 0.307 and a range of 0.173–0.667. Among these compounds, 1 compound (0.19%) was located in the high-similarity region, 3 compounds (0.57%) were located in the intermediate-similarity region, and 518 compounds (99.23%) were located in the low-similarity region. Although the GCN-selected subset exhibited a slightly higher mean $T_{c_{\max}}$ than the overall generated library, the vast majority of selected compounds remained structurally dissimilar to the reference dataset.

These findings indicated low nearest-neighbour similarity to the model-development dataset for most generated compounds, including those retained by GCN screening, under the ECFP4 representation. This analysis described structural coverage but did not establish a validated applicability domain or quantify predictive reliability for the generated compounds. The GCN outputs were therefore used as first-stage computational prioritization scores, not as calibrated probabilities or definitive evidence of VEGFR2 inhibitory activity.

3.4 Molecular docking prioritization and predicted ADMET profiles of candidate compounds

Molecular docking with AutoDock Vina was used to characterize static binding configurations of 522 GCN-selected molecules and the co-crystallized ligand AX4. Candidate Vina-predicted docking scores ranged from -14.32 to -6.06 kcal/mol and showed a right-skewed distribution, with most values between -9.0 and -7.0 kcal/mol. Under the same protocol, AX4 scored -11.56 kcal/mol. Nine MolGPT-generated molecules with more negative scores than AX4 were retained.

Vina scores provide relative rankings only within a specific protocol and scoring function and cannot be equated with experimental binding affinities or true binding free energies [26]. The nine molecules were therefore considered docking-prioritized candidates, not experimentally established high-affinity binders. Ranked by ascending Vina score, they were designated Cpd1–Cpd9 and subjected to binding position analysis and predicted ADMET profiling. Their physicochemical properties, GCN output scores, docking results, and predicted ADMET profiles are summarized in Table 2.

Cpd1 had the most favorable Vina-predicted docking score and a GCN output score of 0.84. This profile supported prioritization, but did not demonstrate strong

VEGFR2 binding. Its synthetic accessibility (SA) score was 5.97, close to the study cutoff of 6. As higher SA scores generally indicate greater predicted synthetic difficulty, Cpd1 may require a more complex synthetic route. The availability of precursors and the feasibility of large-scale preparation were not assessed.

The octanol–water partition coefficient (logP) was evaluated against a Lipinski-based range of 0–5 relevant to oral drug-likeness. The predicted logP values of the nine candidates ranged from 0.98 to 4.38, all within the Lipinski-based reference range of 0–5. Aqueous solubility was assessed using the estimated solubility (ESOL) logS model. Cpd3, Cpd5, Cpd7 and Cpd8 were classified as moderately soluble, with predicted logS values of -5.95 to -4.48 . The others were classified as soluble, with values greater than -4.0 . Cpd4 combined relatively low predicted logP with more favorable predicted aqueous solubility, but this profile did not establish gastrointestinal dissolution, membrane permeability, or oral absorption.

Cytochrome P450 3A4 (CYP3A4), a major hepatic cytochrome P450 (CYP) isoform, contributes to drug metabolism and drug–drug interaction (DDI) risk [36]. Cpd1 and Cpd3 had high predicted probabilities of CYP3A4 inhibition (≥ 0.92), indicating a potential metabolic liability requiring further assessment. Cpd4, Cpd7, Cpd8, and Cpd9 had low predicted CYP3A4 inhibition scores of 0.00–0.02. These values indicated a low predicted likelihood within the model, but did not establish normal hepatic metabolism or the absence of clinically relevant risk of DDI.

Blockade of the human ether-à-go-go-related gene (hERG) potassium channel is associated with drug-induced QT-interval prolongation and potentially fatal ventricular arrhythmias [20]. Cpd4, Cpd6, and Cpd9 had the lowest predicted probabilities of hERG inhibition, all at or below the prespecified cutoff of 0.40. These predictions indicated lower hERG-related liabilities within the model but did not establish cardiac safety. Drug-induced liver injury (DILI) is an important cause of candidate attrition [2], and all candidates had DILI prediction scores of 0.83–1.00.

The same ADMETlab 3.0 analysis was applied to 19 marketed or clinical-stage VEGFR2 inhibitors. Anlotinib scored 0.70 and was classified as intermediate risk, whereas the other agents scored 0.82–1.00. High DILI prediction scores were therefore not unique to the candidates. However, this comparison did not establish high predicted DILI risk as a general property of VEGFR2 inhibitors. Liver-function abnormalities and hepatotoxicity have been reported with several VEGFR2-targeted agents [33], but these observations cannot validate predictions for the present candidates.

ADMETlab DILI outputs represent predicted risk tendencies rather than toxicological conclusions. The results neither demonstrated acceptable safety nor excluded potential hepatotoxicity. Candidate scores overlapped with most comparators without a clearly higher model-predicted signal, but numerical overlap did not establish equivalent clinical risk. DILI liability should remain a priority for structural optimization and experimental safety assessment, not the sole exclusion criterion at this exploratory stage.

The quantitative estimate of drug-likeness (QED) integrates multiple physicochemical properties and empirical desirability distributions derived from approved drugs. All nine candidates had QED values greater than 0.5, indicating moderately favorable

composite profiles without establishing an optimal physicochemical balance. Except for Cpd1, whose SA score approached the study cutoff, the candidates had more favorable predicted synthetic accessibility.

Integration of docking results, GCN output scores, and predicted ADMET profiles prioritized Cpd4, Cpd6, Cpd7, and Cpd9 for similarity matching and exploratory experimental assessment using accessible surrogate compounds. Although Cpd1 and Cpd3 had more favorable Vina-predicted docking scores, both showed high predicted probabilities of CYP3A4 inhibition (≥ 0.92), which reduced their overall priority. Predicted DILI scores were high across most candidates and were therefore treated as warning signals rather than as an exclusion criterion at this exploratory stage. Cpd4, Cpd6, Cpd7, and Cpd9 were retained based on their overall balance of GCN-predicted activity, docking performance, predicted CYP3A4 inhibition, hERG liability, and drug-likeness properties. Similarity searches were subsequently performed to identify experimentally accessible surrogates.

The chemical structures of the four prioritized candidates are shown in Figure 3. Cpd4 contains a 2-methyl-4H-pyrido[1,2-a]pyrimidin-4-one core linked to a 1,4-diazepane ring, with the second nitrogen acylated by a furan-2-carbonyl group. Cpd6 features an N-acetylundoline scaffold connected through a carboxamide linker to an indazole moiety. Cpd7 comprises a trimethyl-substituted quinolin-2(1H)-one framework bearing a pyrazolylphenyl substituent, whereas Cpd9 contains an N-ethylsulfamoyl-substituted benzamide linked through an amide group to a 5-cyanothiazol-2-yl moiety. These four molecular structures were subsequently used as queries for similarity-based matching against commercially available compounds in the TargetMol Natural Product Library.

Table 2 Physicochemical properties, GCN output scores, docking scores, and predicted ADMET profiles of the nine candidates.

Compound	Docking score	GCN score	QED ¹	SA ²	logP	logS	CYP3A4 inhibition	DILI ³	hERG at 10 μ M
Cpd1	-14.32	0.84	0.60	5.97	2.11	-2.71	0.92	1.00 ⁴	0.45
Cpd2	-13.48	0.53	0.53	2.40	2.31	-3.52	0.60	0.83	0.65
Cpd3	-13.39	0.87	0.87	2.38	3.33	-4.60	0.99	1.00 ⁴	0.48
Cpd4	-12.26	0.65	0.71	2.41	0.98	-2.86	0.01	0.98	0.33
Cpd5	-12.22	0.56	0.72	2.22	3.45	-4.48	0.12	0.95	0.61
Cpd6	-11.93	0.87	0.76	2.19	1.64	-3.07	0.19	0.99	0.35
Cpd7	-11.83	0.94	0.57	2.60	3.84	-5.19	0.00	0.99	0.40
Cpd8	-11.79	0.73	0.60	2.13	4.38	-5.95	0.01	1.00 ⁴	0.48
Cpd9	-11.69	0.67	0.86	2.30	1.69	-3.83	0.02	1.00 ⁴	0.38

¹Quantitative estimate of drug-likeness (QED).

²Synthetic accessibility (SA).

³Drug-induced liver injury (DILI).

⁴Probabilistic outputs were rounded to two decimal places; values rounded to 1.00 do not indicate absolute certainty.

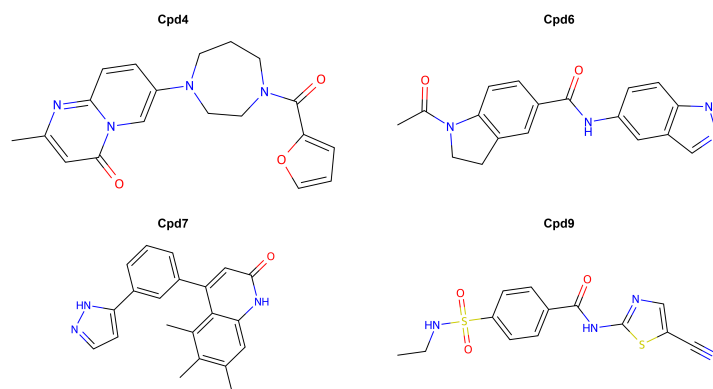


Fig. 3 Two-dimensional chemical structures of the four prioritized MolGPT-generated candidates.

3.5 Similarity-matching results

For the four prioritized MolGPT-generated candidates, Cpd4, Cpd6, Cpd7, and Cpd9, six experimentally accessible similarity-matched surrogate compounds were identified for subsequent exploratory in vitro testing. Cpd4 and Cpd6 were each associated with two surrogate compounds, whereas Cpd7 and Cpd9 were each associated with one.

The Tanimoto coefficients for most generated molecule-surrogate pairs ranged from 0.25 to 0.35. These values indicated low overall ECFP4 fingerprint similarity and substantial differences in their overall molecular structures. Although the selected pairs shared some local fingerprint features, the Tanimoto coefficients alone did not establish conservation of key pharmacophores. In particular, the Tanimoto similarity between levomefolic acid and Cpd9 was only 0.259259.

In a broad exploratory sense, these matches were compatible with a scaffold-hopping rationale in which locally related structural features were examined across different overall scaffolds. However, because the MolGPT-generated candidates had not been experimentally shown to inhibit VEGFR2, these pairs could not be regarded as confirmed scaffold hops. The matching process expanded the range of accessible structures included in exploratory testing but did not establish activity transfer from the generated molecules to the surrogate compounds. Detailed matching results are presented in Table 3.

3.6 In vitro VEGFR2 kinase inhibition assay results

To assess whether the similarity-matched compounds exhibited in vitro inhibition of VEGFR2, six experimentally accessible surrogate compounds were evaluated using a VEGFR2 kinase assay. These compounds included prazosin, trimetozine, metacetamol, paroxypropione, tectoquinone, and levomefolic acid. Levomefolic acid was matched to Cpd9, although their structural similarity was low, with a Tanimoto coefficient of 0.259.

The testing laboratory reported IC_{50} values of $> 40,000$ nM for prazosin, trimetozine, metacetamol, paroxypropione, and tectoquinone, as summarized with

Table 3 Similarity matching between the prioritized MolGPT-generated molecules and experimentally accessible surrogate compounds.

Generated compound	Surrogate compound	Tanimoto coefficient (T_c)
Cpd4	prazosin	0.576923
Cpd4	trimetozine	0.329412
Cpd6	metacetamol	0.351351
Cpd6	paroxypropione	0.257143
Cpd7	tectoquinone	0.322581
Cpd9	levomefolic acid	0.259259

interpretive qualifications in Table ?? . For levomefolic acid, the laboratory reported an apparent relative IC_{50} of 9,038 nM.

However, the concentration–response profile of levomefolic acid was non-monotonic. The maximum observed mean inhibition was $56.6 \pm 5.4\%$ at $20 \mu\text{M}$, decreasing to $12.8 \pm 13.1\%$ at $40 \mu\text{M}$. The duplicate wells also showed greater variability at the highest tested concentration. The cause of this high-concentration deviation was not established by the available experimental records. Accordingly, the apparent IC_{50} was considered an exploratory fitted estimate rather than a reliable quantitative measure of VEGFR2 inhibitory potency. The positive control, staurosporine, showed concentration-dependent inhibition with a fitted IC_{50} of 4.41 nM (Figure 4).

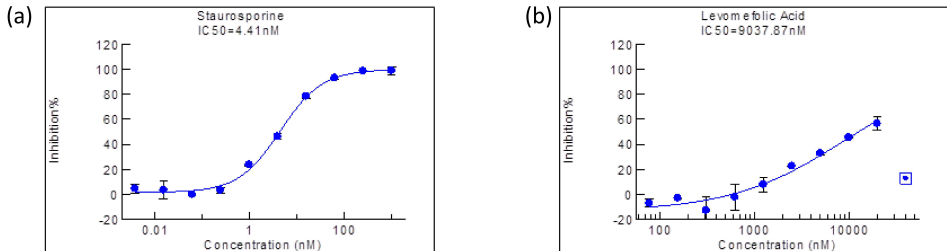


Fig. 4 In vitro VEGFR2 kinase inhibition assay. (a) Dose–response curve of the positive control staurosporine against VEGFR2. (b) Dose–response curve of levomefolic acid against VEGFR2.

The levomefolic acid response remained an exploratory observation and did not establish a reproducible VEGFR2 inhibitory effect. Given its low structural similarity to Cpd9, this response also could not validate Cpd9 activity or activity transfer through similarity matching.

Levomefolic acid was subsequently examined in exploratory computational analyses, with sorafenib included as a reference ligand. Molecular docking and MD simulations were used to examine predicted binding modes and simulated complex dynamics. These analyses were intended to generate structural hypotheses, not to confirm VEGFR2 binding or resolve the uncertainty in the experimental response.

Table 4 VEGFR2 inhibition at the highest tested concentrations and laboratory-reported IC₅₀ results for the six similarity-matched surrogate compounds and the positive control.

Compound	Generated compound	Inhibition (%) ³	Laboratory-reported IC ₅₀ (nM) ²
prazosin	Cpd4	-0.1 ± 2.0	> 40,000
trimetozine	Cpd4	11.2 ± 1.9	> 40,000
metacetamol	Cpd6	5.7 ± 11.0	> 40,000
paroxypropione	Cpd6	4.0 ± 2.6	> 40,000
tectoquinone	Cpd7	10.0 ± 0.3	> 40,000
levomefolic acid	Cpd9	12.8 ± 13.1	9,038 ¹
staurosporine	Positive control	98.94 ± 3.29	4.41

¹The laboratory-reported value for levomefolic acid is an apparent relative IC₅₀ obtained with the upper asymptote fixed at 100%. Mean inhibition peaked at 20 μ M and decreased at 40 μ M; the cause of this deviation was not established. The fitted value is retained as an exploratory result, not as a reliable quantitative estimate of VEGFR2 inhibitory potency.

²IC₅₀ entries reproduce the testing laboratory’s reported results. For trimetozine, metacetamol, and paroxypropione, > 40,000 nM indicates that 50% mean inhibition was not reached over the tested range. For prazosin and tectoquinone, anomalous concentration–response profiles prevent interpreting these entries as established potency bounds.

³Inhibition was measured at 40 μ M for the six test compounds and at 1 μ M for staurosporine. Values are means $\pm$ sample SD from two technical replicate wells per concentration in one assay run.

3.7 Binding Mode Analysis

Among the tested surrogates, levomefolic acid showed a relatively more pronounced experimental response, although it remained limited and uncertain. It was therefore selected for exploratory molecular docking and MD simulations to examine possible binding modes with VEGFR2. The analyses were performed on levomefolic acid, not the corresponding MolGPT-generated molecule Cpd9, and could not provide evidence of Cpd9 binding or activity.

Docking was used to examine predicted poses and interactions within the VEGFR2 active pocket, including hydrogen bonds, hydrophobic contacts, and aromatic-ring-related interactions. Sorafenib contacted Glu885, Lys868, Val848, Ile888, Leu889, Asp1046, Ile1044, Phe1047, and Cys919 and occupied the ATP-binding pocket of the VEGFR2 kinase domain.

Levomefolic acid adopted a distinct predicted pose involving Leu1035, Asn923, Phe1047, Asp1052, and Asp1056 (Fig. 5). Although the two predicted binding modes differed substantially, both involved Phe1047. This shared contact suggested a possible role in ligand recognition, but docking alone could not establish its functional importance. The sorafenib pose more closely resembled the expected pocket occupancy

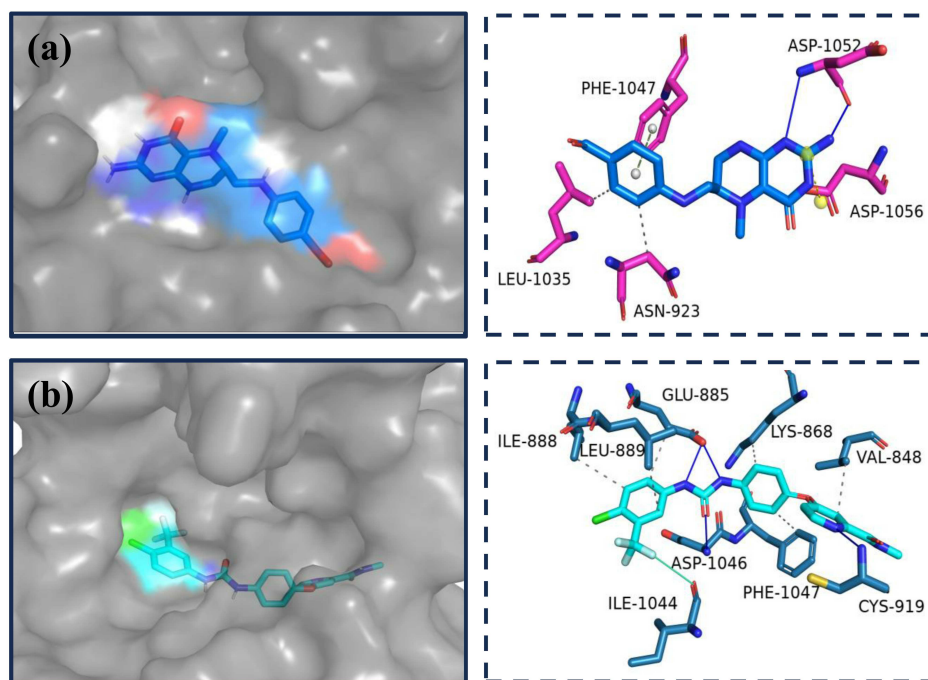


Fig. 5 Predicted binding modes of levomefolic acid and sorafenib in the VEGFR2 binding pocket. (A) Surface binding mode and key residue interactions of levomefolic acid with VEGFR2. (B) Surface binding mode and key residue interactions of sorafenib with VEGFR2.

of a classical VEGFR2 inhibitor. By comparison, the levomefolic acid pose involved several polar residues, including Asn923, Asp1052, and Asp1056, indicating a potentially distinct binding orientation. This pose was used as the starting structure for MD simulations and computational analysis of complex behavior. These docking observations neither demonstrated experimental binding affinity nor confirmed a physical interaction between levomefolic acid and VEGFR2.

3.8 Molecular dynamics simulations and MM–PBSA analyses

3.8.1 Global conformational deviation assessed by RMSD

RMSD characterized time-dependent conformational deviations relative to the selected reference structure. As shown in Fig. 6A, both VEGFR2–ligand complexes initially adjusted away from their docked conformations during the 100-ns simulations.

The sorafenib–VEGFR2 complex maintained comparatively low and stable RMSD values, with a complete-trajectory mean of 1.22 Å and a late-stage value of 1.22 ± 0.18 Å. These values indicated that it reached a relatively stable RMSD range early

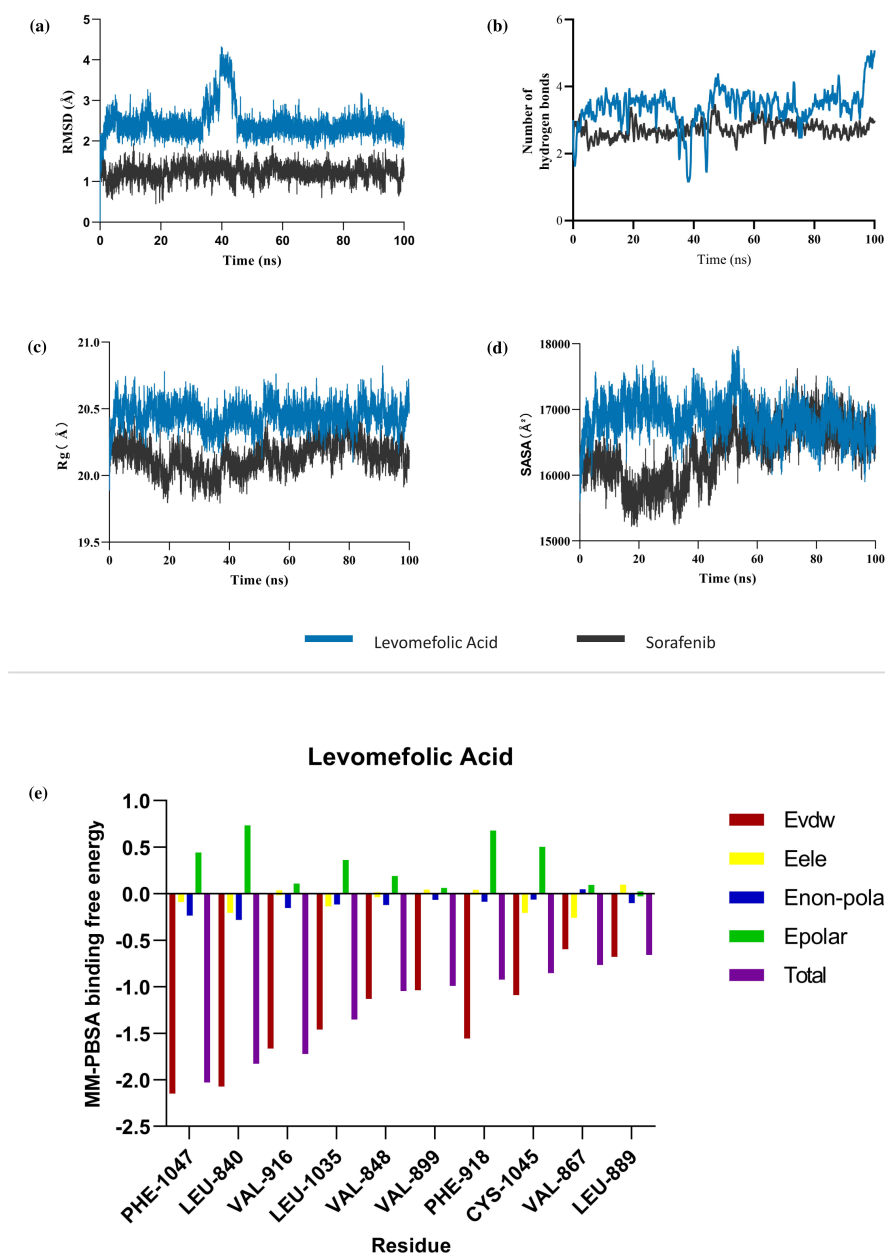


Fig. 6 Molecular dynamics and MM-PBSA analyses of levomefolic acid-VEGFR2 and sorafenib-VEGFR2 complexes. (A) RMSD profiles during the 100 ns molecular dynamics simulation. (B) Number of hydrogen bonds formed during the simulation. (C) Radius of gyration (R_g) profiles. (D) Solvent-accessible surface area (SASA) profiles. (E) Per-residue MM-PBSA energy decomposition of the levomefolic acid-VEGFR2 complex.

in the simulation. The levomefolic acid–VEGFR2 complex showed larger deviations, with an overall mean of 2.43 Å and a late-stage value of 2.34 ± 0.18 Å.

A pronounced transient fluctuation occurred in the levomefolic acid trajectory at approximately 35–45 ns, when RMSD exceeded 4 Å. This pattern was consistent with transient conformational rearrangement, although RMSD alone could not distinguish local protein rearrangement from ligand reorientation or a combination of both. RMSD subsequently decreased and remained at approximately 2.0–2.5 Å without a sustained upward trend.

Overall, the sorafenib complex showed smaller and less variable conformational deviations. The levomefolic acid complex underwent a larger transient deviation but subsequently maintained a comparatively stable late-stage RMSD range. These observations describe simulated conformational dynamics and do not independently demonstrate stable ligand binding or experimental complex stability.

3.8.2 Hydrogen-bond dynamics

Protein–ligand hydrogen bonds were monitored to characterize polar contacts within the simulated complexes. Both complexes formed hydrogen bonds over 100 ns (Fig. 6B), with mean counts of 3.41 for levomefolic acid and 2.73 for sorafenib.

The levomefolic acid–VEGFR2 complex generally maintained two to five hydrogen bonds, increasing to six or seven during several late-stage intervals. The sorafenib complex generally formed approximately one to three hydrogen bonds and reached four at only a few time points. Levomefolic acid therefore sampled a more extensive hydrogen-bond network within the simulated trajectory.

Although the hydrogen-bond count for levomefolic acid fluctuated, it remained generally higher than that for sorafenib, indicating recurrent polar contacts within the simulated pocket. However, hydrogen-bond counts did not establish the persistence of individual interactions or greater binding stability. Time-resolved occupancy and correlation analyses would be required to determine whether these contacts compensated for the larger RMSD fluctuations. The lower count for sorafenib was consistent with possible contributions from hydrophobic, van der Waals, or other nonpolar interactions but did not establish their relative importance.

3.8.3 Global compactness and solvent exposure

Rg and SASA were analyzed to characterize global compactness and solvent exposure. The Rg of the levomefolic acid–VEGFR2 complex remained within a narrow range over 100 ns, with a mean of 20.45 ± 0.09 Å (Fig. 6C). The absence of a sustained increase or decrease indicated no progressive global expansion or collapse within the trajectory.

The levomefolic acid complex had a slightly higher Rg than the sorafenib complex, corresponding to a somewhat less compact simulated conformation. However, the absence of a progressive Rg change indicated no continuing structural expansion.

SASA fluctuated dynamically without a sustained increase and had a mean of $16,866 \pm 287$ Å² (Fig. 6D), indicating relatively consistent solvent exposure. Neither the Rg nor SASA profile independently established stable ligand binding or experimental complex stability.

3.8.4 MM-PBSA-estimated binding energetics

MM-PBSA calculations estimated the relative binding energetics of both complexes using 101 snapshots from the 90–100 ns interval of each trajectory (Table 5). Within the same protocol, a more negative estimated ΔG_{bind} indicates more favorable calculated binding energetics. These estimates cannot be equated with experimental affinity or true thermodynamic binding free energy.

Both complexes had negative estimated ΔG_{bind} values. The levomefolic acid-VEGFR2 complex yielded -14.244 ± 3.435 kcal/mol, whereas the sorafenib-VEGFR2 complex yielded -35.573 ± 2.782 kcal/mol (mean $\pm$ sample SD across the analyzed snapshots). Sorafenib therefore showed more favorable estimated binding energetics under the applied protocol. This relative result was consistent with its use as an established VEGFR2 inhibitor and reference ligand but did not validate the MM-PBSA estimates.

Energy decomposition showed favorable contributions from van der Waals, electrostatic, and nonpolar solvation terms in both complexes, whereas polar solvation opposed binding. Sorafenib had a more favorable van der Waals contribution and a smaller polar solvation penalty, contributing to its more negative estimated ΔG_{bind} .

Levomefolic acid had a more favorable electrostatic contribution, consistent with its larger sampled hydrogen-bond count, although these analyses did not establish causality. Its comparatively large polar solvation penalty offset much of the favorable electrostatic, van der Waals, and nonpolar solvation contributions, producing a less favorable total estimate than sorafenib.

Overall, the two complexes showed different balances of estimated energy components. Sorafenib benefited mainly from more favorable van der Waals interactions and a smaller polar solvation penalty. Levomefolic acid showed a favorable electrostatic contribution that was substantially offset by polar solvation. These estimates support comparative mechanistic hypotheses but do not demonstrate stable VEGFR2 binding, experimental affinity, or thermodynamic stability for levomefolic acid.

Table 5 MM-PBSA binding-energy estimates and their components for the levomefolic acid-VEGFR2 and sorafenib-VEGFR2 complexes. Values are means $\pm$ sample SD across 101 snapshots from 90–100 ns, expressed in kcal/mol. SD describes within-trajectory variation, not variability across independent simulations.

Compound	Van der Waals	Electrostatic	Polar solvation	Nonpolar solvation	ΔG_{bind}
Levomefolic acid	-49.763 ± 3.339	-17.756 ± 4.514	58.585 ± 3.878	-5.310 ± 0.214	-14.244 ± 3.435
Sorafenib	-59.594 ± 2.551	-15.202 ± 2.308	44.814 ± 2.459	-5.591 ± 0.182	-35.573 ± 2.782

3.8.5 Per-residue MM-PBSA energy decomposition

Per-residue MM-PBSA decomposition identified residues with the largest estimated energetic contributions in the simulated levomefolic acid-VEGFR2 complex. The 10

residues with the most favorable total contributions are shown in Fig. 6E. Phe1047 had the largest favorable contribution at -2.028 kcal/mol, followed by Leu840, Val916, Leu1035, Val848, Val899, Phe918, Cys1045, Val867, and Leu889. Their negative values indicated favorable relative contributions under the applied decomposition protocol.

Phe1047 and Leu1035 appeared in both the docking and decomposition analyses. Phe1047 had the largest favorable decomposed contribution and was therefore identified as a residue of interest for further investigation, while Leu1035 also contributed comparatively favorably. Several hydrophobic residues, including Val848, Val899, Leu840, and Phe918, also had negative contributions, consistent with contributions from hydrophobic contacts and van der Waals interactions within the simulated complex.

The favorable contributions of the 10 residues arose predominantly from the van der Waals term. In particular, Phe1047, Leu840, Val916, and Leu1035 showed comparatively favorable negative ΔE_{vdW} values. In contrast, polar solvation contributions were positive for most residues and partially offset favorable van der Waals, electrostatic, and nonpolar solvation terms. These estimates characterize only the simulated energy decomposition and do not establish experimentally confirmed binding hotspots or residue-level determinants of VEGFR2 recognition.

4 Discussion

VEGFR2 is a key mediator of tumor angiogenesis and an established target for antiangiogenic therapy. However, current inhibitors, including sorafenib and sunitinib, remain limited by insufficient selectivity, dose-limiting toxicities, pharmacokinetic liabilities, and acquired resistance. Therefore, structurally novel VEGFR2 inhibitor-like molecules with potential for further optimization remain relevant for early anticancer drug discovery.

We developed a multistage workflow that integrates molecular generation using the MolGPT model, GCN-based activity prediction, molecular docking, ADMET assessment, similarity matching, exploratory in vitro kinase testing, and MD simulations. The workflow was designed to address two practical challenges: first, to explore chemical space beyond conventional compound libraries through generative AI; and second, to bridge computationally prioritized but experimentally inaccessible molecules with commercially available surrogate compounds. This approach extends candidate exploration beyond our previous work, which depended on screening existing compound databases, by using deep learning to generate new structures and exploring similarity matching as a practical connection between generated molecules and experimentally accessible compounds. Accordingly, the present study should be interpreted primarily as an exploratory evaluation of a computational-to-experimental prioritization workflow rather than as experimental validation of newly generated VEGFR2 inhibitors.

The Transformer decoder-based MolGPT model generated 4,992 unique molecules with high validity, uniqueness, novelty, and internal diversity under the evaluated MOSES metrics. Unlike conventional virtual screening restricted to existing libraries,

generative modeling enabled exploration of molecular structures beyond the compounds available in the screening database. However, MolGPT was not specifically fine-tuned on known VEGFR2 inhibitors; therefore, target relevance was introduced primarily through downstream computational screening. A previously evaluated GCN model trained on ChEMBL VEGFR2 data served as a prescreening tool, retaining 522 candidates from the generated molecules. Molecular docking then identified nine candidates with more favorable Vina-predicted scores than the co-crystallized ligand AX4, and ADMET profiling was used to assess physicochemical properties, metabolic liabilities, and toxicity-related risks. This multistage computational process was intended to reduce false-positive prioritization; however, none of these computational filters established biological activity.

The similarity-based chemical-space coverage analysis provided context for interpreting the GCN-based prioritization of the MolGPT-generated compounds. Most generated molecules, including the majority of compounds retained by GCN-based screening, showed low structural similarity to the VEGFR2-related compounds in the original model-development dataset. This finding indicates that MolGPT explored chemical space substantially different from that represented by the reference dataset, consistent with the intended generation of structurally novel candidate molecules. At the same time, the limited structural overlap suggests that GCN predictions for many generated compounds involved structural extrapolation beyond the chemical space represented by the model-development dataset. Consequently, the GCN outputs were interpreted primarily as exploratory prioritization scores rather than as calibrated probabilities or definitive evidence of VEGFR2 inhibitory activity. This consideration supported the use of GCN as an initial filtering step, followed by molecular docking and, for levomefolic acid, subsequent molecular dynamics simulations. Nevertheless, the limited structural overlap with the model-development chemical space remains an important limitation, and direct experimental testing is ultimately required to establish the VEGFR2 inhibitory activity of the generated compounds.

A practical limitation was that the prioritized generated molecules were not commercially available and were considered difficult to access rapidly through direct synthesis. Therefore, structurally related commercial compounds were identified as surrogates for exploratory experimental assessment. This similarity-based bridging strategy did not constitute direct testing of the generated molecules, and the generally limited structural similarity between the surrogates and their corresponding generated molecules represents an important limitation of this approach.

Exploratory testing of the six similarity-matched surrogates did not establish a reproducible VEGFR2 inhibitory effect. For levomefolic acid, the laboratory-reported apparent relative IC_{50} of 9,038 nM was limited by a non-monotonic concentration-response profile and remained an exploratory fitted estimate. The cause of the response decrease at the highest concentration was not established. Moreover, the low Tanimoto similarity of 0.259 between levomefolic acid and Cpd9 further limited extrapolation to the generated molecule. These findings therefore did not validate Cpd9 activity or activity transfer through the similarity-matching strategy.

The limited experimental yield also illustrates an important limitation of two-dimensional similarity matching. ECFP4-based Tanimoto similarity does not necessarily preserve three-dimensional molecular shape, pharmacophore geometry, conformational preferences, or protein–ligand interaction patterns required for kinase inhibition. Subsequent molecular docking and MD simulations were therefore used only to generate structural hypotheses concerning the possible behavior of levomefolic acid in the VEGFR2 binding pocket. Levomefolic acid showed distinct predicted interactions involving Asn923, Asp1052, and Asp1056, whereas sorafenib engaged canonical ATP-pocket residues including Glu885, Lys868, and Asp1046. MM-PBSA calculations suggested that favorable electrostatic contributions for levomefolic acid were offset by larger polar solvation penalties, whereas sorafenib exhibited more favorable estimated binding energetics. These computational analyses may inform hypotheses for future compound design, but they cannot confirm experimental binding, establish VEGFR2 inhibitory activity, or explain the limited in vitro response. Orthogonal experimental approaches, including direct binding assays and, where feasible, structural characterization, would be required to validate these computational hypotheses.

Several methodological limitations constrain the conclusions that can be drawn from the present study. Most importantly, none of the prioritized MolGPT-generated molecules, including Cpd4, Cpd6, Cpd7, and Cpd9, were synthesized or tested directly in the VEGFR2 kinase assay. The present data therefore do not experimentally validate these molecules as VEGFR2 inhibitors. Although a similarity-based chemical-space coverage analysis was performed, it characterized fingerprint-based structural proximity rather than a validated predictive applicability domain. The predefined similarity regions were not calibrated against GCN performance using an independent labelled dataset, nor did the analysis directly characterize the representation space learned by the GCN. Prediction reliability therefore remains uncertain for generated compounds without close analogues in the model-development dataset. Although the co-crystallized ligand AX4 was included as a reference in the docking calculations, the docking protocol was not independently validated by re-docking AX4 and calculating the heavy-atom RMSD relative to its crystallographic pose. The docking results should therefore be interpreted as comparative computational rankings and structural hypotheses rather than as validated predictions of binding affinity or pose accuracy. Similarly, the MM-PBSA binding-energy estimates are model-dependent computational quantities and cannot substitute for direct binding or functional assays. Furthermore, the similarity-based bridging strategy relied primarily on two-dimensional Tanimoto similarity, and only six surrogate compounds were experimentally tested. The limited experimental evidence does not allow the effects of surrogate selection and structural dissimilarity to be distinguished from the intrinsic activity of the generated candidates.

Overall, this study explores an AI-assisted framework integrating MolGPT-based molecular generation, GCN-based prioritization, molecular docking, predicted ADMET assessment, similarity matching, exploratory surrogate testing, and molecular simulation. The principal contribution is methodological rather than the identification or validation of a new VEGFR2 inhibitor. The current experimental evidence does not

establish VEGFR2 inhibitory activity for any of the prioritized MolGPT-generated molecules and does not validate activity transfer through the similarity-matching strategy. Future studies should therefore prioritize direct synthesis and biological testing of generated candidates, evaluation of closer structural analogues, incorporation of three-dimensional shape and pharmacophore similarity, empirical calibration and validation of the GCN predictive applicability domain, and orthogonal experimental validation of predicted binding interactions.

Declarations

Author contributions. Taiying Li: Conceptualization, Methodology, Investigation, Visualization, and Writing—original draft. Shuning Diao: Investigation, Validation, and Visualization. Yuxiang He: Investigation, Validation, and Visualization. Shengzhen Hou: Validation and Visualization. Jinping Zhang: Conceptualization, Project administration, Supervision, Resources, Software, and Writing—review and editing. All authors approved the final version of the manuscript.

Funding. This research received no external funding.

Conflict of interest. The authors declare that they have no competing interests.

Data availability. The raw data supporting the conclusions of this article will be made available by the authors without reservations.

References

- [1] Abraham MJ, Murtola T, Schulz R, et al (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- [2] Allison R, Guraka A, Shawa IT, et al (2023) Drug induced liver injury - a 2023 update. *J Toxicol Environ Health B Crit Rev* 26(8):442–467. <https://doi.org/10.1080/10937404.2023.2261848>
- [3] Arshia AH, Shadravan S, Solhjoo A, et al (2021) De novo design of novel protease inhibitor candidates in the treatment of sars-cov-2 using deep learning, docking, and molecular dynamic simulations. *Comput Biol Med* 139:104967. <https://doi.org/10.1016/j.combiomed.2021.104967>
- [4] Bagal V, Aggarwal R, Vinod PK, et al (2022) Molgpt: Molecular generation using a transformer-decoder model. *J Chem Inf Model* 62(9):2064–2076. <https://doi.org/10.1021/acs.jcim.1c00600>
- [5] Bodun DS, Ibrahim IO, Bashiru M, et al (2025) De novo design and bioactivity prediction of mitotic kinesin eg5 inhibitors using mpnn and lstm-based transfer learning. *Comput Biol Med* 195:110672. <https://doi.org/10.1016/j.combiomed.2025.110672>

- [6] Bray F, Laversanne M, Sung H, et al (2024) Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 74(3):229–263. <https://doi.org/10.3322/caac.21834>
- [7] Carmeliet P, Jain RK (2000) Angiogenesis in cancer and other diseases. *Nature* 407(6801):249–257. <https://doi.org/10.1038/35025220>
- [8] Chakraborty A, Krishnan V, Thamotharan S (2025) Generative adversarial network (gan) model-based design of potent sars-cov-2 m(pro) inhibitors using the electron density of ligands and 3d binding pockets: insights from molecular docking, dynamics simulation, and mm-gbsa analysis. *Mol Divers* 29(4):3059–3075. <https://doi.org/10.1007/s11030-024-11047-9>
- [9] Eberhardt J, Santos-Martins D, Tillack AF, et al (2021) Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *J Chem Inf Model* 61(8):3891–3898. <https://doi.org/10.1021/acs.jcim.1c00203>
- [10] Eggert A, Ikegaki N, Kwiatkowski J, et al (2000) High-level expression of angiogenic factors is associated with advanced tumor stage in human neuroblastomas. *Clin Cancer Res* 6(5):1900–8
- [11] Fu L, Shi S, Yi J, et al (2024) Admetlab 3.0: an updated comprehensive online admet prediction platform enhanced with broader coverage, improved performance, api functionality and decision support. *Nucleic Acids Res* 52(W1):W422–w431. <https://doi.org/10.1093/nar/gkae236>
- [12] Genheden S, Ryde U (2015) The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10(5):449–61. <https://doi.org/10.1517/17460441.2015.1032936>
- [13] Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646–74. <https://doi.org/10.1016/j.cell.2011.02.013>
- [14] Hou S, Diao S, He Y, et al (2025) Identification of potent inhibitors of potential vegfr2: a graph neural network-based virtual screening and in vitro study. *J Enzyme Inhib Med Chem* 40(1):2518192. <https://doi.org/10.1080/14756366.2025.2518192>
- [15] Kovar O, Kufa M, Finger V, et al (2025) Scaffold hopping in tuberculosis drug discovery: Principles, applications, and case studies. *J Med Chem* 68(20):20903–20929. <https://doi.org/10.1021/acs.jmedchem.5c01100>
- [16] Kumari R, Kumar R, Lynn A (2014) g_mmpbsa—a gromacs tool for high-throughput mm-pbsa calculations. *J Chem Inf Model* 54(7):1951–62. <https://doi.org/10.1021/ci500020m>

- [17] Lipinski CA, Lombardo F, Dominy BW, et al (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46(1-3):3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
- [18] Liu YZ, Wang XL, Wang XY, et al (2016) De novo design of vegfr-2 tyrosine kinase inhibitors based on a linked-fragment approach. *J Mol Model* 22(9):222. <https://doi.org/10.1007/s00894-016-3088-8>
- [19] Liu ZL, Chen HH, Zheng LL, et al (2023) Angiogenic signaling pathways and anti-angiogenic therapy for cancer. *Signal Transduct Target Ther* 8(1):198. <https://doi.org/10.1038/s41392-023-01460-1>
- [20] Lopez-Medina AI, Chahal CAA, Luzum JA (2022) The genetics of drug-induced qt prolongation: evaluating the evidence for pharmacodynamic variants. *Pharmacogenomics* 23(9):543–557. <https://doi.org/10.2217/pgs-2022-0027>
- [21] McInnes L, Healy J, Saul N, et al (2018) Umap: Uniform manifold approximation and projection. *Journal of Open Source Software* 3(29). <https://doi.org/10.21105/joss.00861>
- [22] Moradi M, Mousavi A, Emamgholipour Z, et al (2023) Quinazoline-based vegfr-2 inhibitors as potential anti-angiogenic agents: A contemporary perspective of sar and molecular docking studies. *Eur J Med Chem* 259:115626. <https://doi.org/10.1016/j.ejmech.2023.115626>
- [23] Morris GM, Huey R, Lindstrom W, et al (2009) Autodock4 and autodock-tools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30(16):2785–91. <https://doi.org/10.1002/jcc.21256>
- [24] Mross K, Frost A, Steinbild S, et al (2012) A phase i dose–escalation study of regorafenib (bay 73–4506), an inhibitor of oncogenic, angiogenic, and stromal kinases, in patients with advanced solid tumors. *Clinical Cancer Research* 18(9):2658–2667. <https://doi.org/10.1158/1078-0432.Ccr-11-1900>
- [25] Nayarisseri A, Abdalla M, Joshi I, et al (2024) Potential inhibitors of vegfr1, vegfr2, and vegfr3 developed through deep learning for the treatment of cervical cancer. *Sci Rep* 14(1):13251. <https://doi.org/10.1038/s41598-024-63762-w>
- [26] Paggi JM, Pandit A, Dror RO (2024) The art and science of molecular docking. *Annu Rev Biochem* 93(1):389–410. <https://doi.org/10.1146/annurev-biochem-030222-120000>
- [27] Polykovskiy D, Zhebrak A, Sanchez-Lengeling B, et al (2020) Molecular sets (moses): A benchmarking platform for molecular generation models. *Front Pharmacol* 11:565644. <https://doi.org/10.3389/fphar.2020.565644>

- [28] Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50(5):742–54. <https://doi.org/10.1021/ci100050t>
- [29] Rogers DJ, Tanimoto TT (1960) A computer program for classifying plants. *Science* 132(3434):1115–8. <https://doi.org/10.1126/science.132.3434.1115>
- [30] Steeghs N, Hovens MM, Rabelink AJ, et al (2006) Vegfr2 blockade in patients with solid tumors: Mechanism of hypertension and effects on vascular function. *Journal of Clinical Oncology* 24(18 Suppl):3037–3037. https://doi.org/10.1200/jco.2006.24.18_suppl.3037
- [31] Vamathevan J, Clark D, Czodrowski P, et al (2019) Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* 18(6):463–477. <https://doi.org/10.1038/s41573-019-0024-5>
- [32] Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., p 6000–6010
- [33] Viganò M, La Milia M, Grassini MV, et al (2023) Hepatotoxicity of small molecule protein kinase inhibitors for cancer. *Cancers (Basel)* 15(6). <https://doi.org/10.3390/cancers15061766>
- [34] Wu C, He T (2024) A survey of applications of vision transformer and its variants. In: *2024 10th IEEE International Conference on Intelligent Data and Security (IDS)*, pp 21–25, <https://doi.org/10.1109/ids62739.2024.00011>
- [35] Yang J, Cai Y, Zhao K, et al (2022) Concepts and applications of chemical fingerprint for hit and lead screening. *Drug Discov Today* 27(11):103356. <https://doi.org/10.1016/j.drudis.2022.103356>
- [36] Zhang Y, Wang Z, Wang Y, et al (2024) Cyp3a4 and cyp3a5: the crucial roles in clinical drug metabolism and the significant implications of genetic polymorphisms. *PeerJ* 12:e18636. <https://doi.org/10.7717/peerj.18636>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData.xlsx](#)