

Generative AI–Driven Design and Experimental Validation of EGFPAN_EMBS, a Novel EGFR Inhibitor in Pancreatic Cancer

Anuraj Nayarisseri

anuraj@eminentbio.com

Eminent Biosciences

Jyothi Kaparapu

GITAM Deemed to be University

Sajal Suhane

The University of Texas at Dallas

Vaishnavi Jadhav

Eminent Biosciences

Yug Shree

Eminent Biosciences

Dhruvi Jungi

Eminent Biosciences

Rashmi Sharma

Shri Vaishnav Vidhyapeeth Vishwavidyalaya

Akansha Sankhala

Eminent Biosciences

Radhika Swami

Eminent Biosciences

Umesh Panwar

AiDD Lab

Murali Aarthi

Eminent Biosciences

Maddala Madhavi

Nizam College, Osmania University

Srinivas Bandaru

Koneru Lakshmiah Educational Foundation (KLEF)

Pranoti Belapurkar

Acropolis Institute of Management Studies and Research: Indore

Hafiz Ahmad

RAK College of Medical Sciences, RAK Medical & Health Sciences University

Keun Woo Lee

AiDD Lab

Francisco Jaime Bezerra Mendonça Junior

State University of Paraíba

Marcus T Scotti

Federal Paraíba University (UFPB)

Luciana Scotti

Federal Paraíba University (UFPB)

Article

Keywords: Generative AI, EGFR inhibitors & Pancreatic cancer, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Reinforcement Learning (RL), MD Simulation, Drug Discovery

Posted Date: July 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10394274/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 28th, 2026. See the published version at <https://doi.org/10.1038/s41598-026-73341-w>.

Abstract

Pancreatic cancer remains a highly lethal malignancy with limited targeted therapeutic options, necessitating innovative drug discovery strategies. The epidermal growth factor receptor (EGFR) is a key driver of tumor progression and a validated therapeutic target. Here, we report an integrated generative artificial intelligence (AI)-driven framework for the discovery of novel EGFR inhibitors, leading to the identification of EGFPAN_EMBS as a chemically distinct and potent candidate. A cannabigerol-inspired scaffold was used as the initial seed, and deep generative models, including variational autoencoders, generative adversarial networks, and reinforcement learning, were employed to explore a chemical space exceeding 1.7 million compounds. Lead prioritization based on drug-likeness, pharmacokinetic properties, and target specificity identified EGFPAN_EMBS as a top candidate. Docking studies revealed improved binding affinity compared to cannabigerol, supported by a well-defined interaction network involving key residues within the ATP-binding pocket. Molecular dynamics simulations demonstrated enhanced structural stability, reduced conformational fluctuations, and persistent intermolecular interactions. Free energy landscape and dynamical cross-correlation analyses indicated confinement to a low-energy conformational basin and reduced long-range correlated motions. Binding free energy calculations further confirmed stronger interaction stability for EGFPAN_EMBS. Experimental validation showed potent nanomolar inhibition of EGFR kinase activity, favorable binding kinetics, and strong antiproliferative effects across pancreatic cancer cell lines with high selectivity. Mechanistically, EGFPAN_EMBS suppressed EGFR autophosphorylation and downstream PI3K–AKT–MAPK signaling, inducing apoptosis and G₁-phase cell-cycle arrest. Collectively, EGFPAN_EMBS represents a promising ATP-competitive EGFR inhibitor and highlights the potential of AI-driven discovery pipelines for next-generation anticancer therapeutics. The compound EGFPAN_EMBS has been deposited in the NCBI PubChem database under Compound ID 10972272 and Substance ID 528216732.

1. Introduction

Pancreatic cancer, predominantly pancreatic ductal adenocarcinoma (PDAC), remains one of the most aggressive malignancies and a leading cause of cancer-related mortality worldwide. Its poor prognosis is driven by late-stage diagnosis, rapid disease progression, and profound resistance to existing therapies. Although surgical resection offers the only curative option, most patients relapse, highlighting the urgent need for effective targeted interventions. PDAC exhibits substantial molecular heterogeneity, which contributes to therapeutic resistance and limits the success of current treatment strategies. Despite extensive efforts to target oncogenic signaling pathways, clinical outcomes remain poor, underscoring the need for more precise and effective therapeutic approaches. Among potential targets, the epidermal growth factor receptor (EGFR) has emerged as a compelling candidate in PDAC(1–4).

EGFR, a receptor tyrosine kinase, regulates cellular proliferation, survival, and invasion, and its dysregulation drives tumor progression through activation of RAS–RAF–MEK–ERK and PI3K–AKT signaling pathways(5–7). In addition, EGFR has been implicated in early tumorigenic events and aggressive disease phenotypes, supporting its continued relevance as a therapeutic target in pancreatic

cancer(8). However, the development of selective and effective EGFR inhibitors for PDAC remains limited, highlighting a critical gap in current therapeutic strategies(9,10). Conventional drug discovery approaches, including structure-based design and high-throughput screening, are constrained by limited chemical space exploration, high cost, and prolonged development timelines. Moreover, these methods rely predominantly on static structural representations and often fail to capture the conformational flexibility of ligands and the dynamic nature of protein–ligand interactions, contributing to poor translation from computational prediction to experimental validation(11–14).

Recent advances in generative artificial intelligence (AI) have enabled the de novo design of structurally diverse molecules by learning latent chemical representations. Approaches such as variational autoencoders, generative adversarial networks, and reinforcement learning frameworks have demonstrated the ability to expand chemical space and optimize molecular properties. However, most existing AI-driven pipelines remain limited by insufficient integration with dynamic structural evaluation and continue to rely heavily on static docking methods, reducing their ability to accurately predict binding stability and functional efficacy. Notably, the application of such integrated generative AI frameworks to EGFR-targeted drug discovery in pancreatic cancer remains largely unexplored(15–19).

Unlike conventional drug discovery pipelines that rely on static structure-based screening or limited chemical space exploration, the present study introduces an integrated, conformer-aware generative AI framework that combines variational autoencoders, generative adversarial networks, and reinforcement learning with dynamic structural validation. This approach enables the generation of chemically diverse and syntactically robust molecular candidates while simultaneously optimizing for target-specific binding and drug-like properties. Importantly, the framework incorporates conformational diversity at the initial stage and extends beyond traditional docking by integrating molecular dynamics simulations, free energy landscape analysis, and dynamical cross-correlation evaluation to capture the dynamic behavior of protein–ligand interactions. By bridging generative modeling with physics-based validation and experimental evaluation, this work establishes a scalable and biologically informed bioinformatics pipeline for de novo drug discovery targeting EGFR in pancreatic cancer.

To address these challenges, we developed an integrated computational–experimental workflow that combines generative AI with deep learning–guided docking and molecular dynamics simulations to capture both structural and dynamic determinants of ligand binding. Starting from a cannabigerol-derived scaffold, this approach enables efficient exploration and prioritization of chemically diverse candidates beyond conventional libraries. Using this strategy, we identified a quinazolinone-based compound, EGFPAN_EMBS, as a potent and selective EGFR inhibitor(Fig. 1). Experimental validation demonstrated nanomolar kinase inhibition, strong antiproliferative activity, and selective targeting of pancreatic cancer cells. Collectively, this work establishes a scalable AI-driven discovery paradigm and highlights its application in identifying novel EGFR-targeted therapeutics for pancreatic cancer.

2. Methods

2.1. Molecular Docking

A set of reported epidermal growth factor receptor (EGFR) inhibitors, including synthetic compounds and bioactive phytochemicals, was curated from the literature to establish a reference dataset for computational screening. Representative molecules were selected for docking analysis (**Tables SM1 & 2, Supplementary Methods**). Ligand structures were prepared using the Open Babel package, including geometry optimization and format conversion. Molecular docking was performed using Gnina to predict binding orientations and interaction profiles within EGFR (20–22). A blind docking protocol was employed to enable unbiased exploration of potential binding regions across the protein surface. For each ligand–protein complex, the top-ranked pose according to the scoring function was selected and analyzed for key intermolecular interactions. Detailed procedures for protein and ligand preparation, docking parameters, and interaction profiling are provided in **Supplementary Methods** (23–27)

2.2. Generative Modelling and Design of Novel EGFR Inhibitors

Novel EGFR-targeting compounds were generated using an integrated deep generative framework combining conditional generative adversarial networks (ConfGAN), variational autoencoders (VAE), and reinforcement learning (RL). The framework was designed to explore EGFR-relevant chemical space and enable the identification of structurally novel inhibitors. ConfGAN was initially employed to generate conformational variants of a cannabigerol-derived scaffold, capturing conformational diversity and key structural features. These conformers were subsequently used as input for the VAE–GAN–RL framework to enable de novo compound generation and optimization(28–31). The integrated architecture enabled exploration of chemically diverse candidates while incorporating conformational and physicochemical constraints derived from known EGFR inhibitors. Compound generation was guided by a multi-parameter reward function integrating a quantitative estimate of drug-likeness (QED), predicted EGFR binding affinity, Lipinski's Rule of Five, and synthetic accessibility analyses(24,27,32–35).

Approximately 1.7 million candidate molecules were generated and prioritized based on structural diversity, predicted pharmacological properties, and model-derived scoring functions. Top candidates were selected for downstream molecular docking and molecular dynamics simulations (Fig. 2). Detailed model architecture and training procedures are described in the **Supplementary Methods**(36–39).

To further strengthen the methodological foundation of the proposed framework, we explicitly incorporated advanced molecular representation and deep learning strategies tailored for bioinformatics-driven drug design. Molecular structures were encoded using SELFIES (Self-Referencing Embedded Strings), which provide a robust and chemically valid representation compared to conventional SMILES, thereby ensuring the generation of syntactically and chemically consistent molecules. The encoded representations were used to train a variational autoencoder (VAE) to learn a continuous latent space of EGFR-relevant chemical features.

The generative architecture integrates a multi-stage pipeline combining VAE, generative adversarial networks (GANs), and reinforcement learning (RL). The VAE component captures latent chemical distributions, while the GAN module enhances structural diversity through adversarial training. Reinforcement learning was subsequently applied to optimize generated molecules using a multi-objective reward function incorporating quantitative estimate of drug-likeness (QED), predicted EGFR binding affinity, Lipinski's Rule of Five, and synthetic accessibility.

The model was trained on a curated dataset of EGFR inhibitors and chemically relevant scaffolds, with training performed over multiple epochs to ensure convergence of latent representations and stability of generated outputs. Model validation was conducted using standard cheminformatics metrics, including validity, novelty, and diversity, along with target-specific docking-based scoring. This integrated bioinformatics framework enables efficient exploration of chemical space while maintaining biological relevance and target specificity, thereby addressing key limitations of traditional structure-based drug discovery approaches.

2.3. Molecular Dynamics Simulation and Conformational Analysis

Molecular dynamics (MD) simulations were performed using GROMACS to evaluate the stability and conformational dynamics of EGFR–ligand complexes. Systems were solvated in an explicit aqueous environment, neutralized with counterions, and subjected to energy minimization followed by equilibration under NVT and NPT conditions. Production simulations were conducted for 200 ns at 300 K and 1 bar. Long-range electrostatic interactions were treated using the Particle Mesh Ewald method, and bond constraints were applied using the LINCS algorithm. Trajectory analyses were performed to characterize structural stability and dynamic behavior. Root-mean-square deviation (RMSD), radius of gyration, hydrogen bonding, and solvent-accessible surface area were evaluated to assess protein stability and ligand binding persistence(24,25,27,32,40–42).

Principal component analysis (PCA) was performed to capture dominant collective motions, while dynamic cross-correlation matrix (DCCM) analysis was used to evaluate correlated and anti-correlated residue motions within the kinase domain. Binding free energies were estimated using the molecular mechanics/generalized Born surface area (MM/GBSA) approach to assess interaction stability. Free energy landscapes (FELs) were constructed based on principal components to identify energetically favorable conformational states sampled during the simulations(43–46). Detailed simulation parameters and analysis protocols are provided in the **Supplementary Methods**. Pharmacokinetic properties and drug-likeness of the selected compounds using in silico ADMET prediction tools are provided in **SupplementaryMethods**(26,47–51).

2.4. Chemical Synthesis of EGFPAN_EMBS

EGFPAN_EMBS was synthesized via a multistep synthetic route starting from commercially available precursors. Key transformations included protection of phenolic groups, preparation of a long-chain

diacid chloride, Friedel–Crafts acylation, reductive methylenation, and final deprotection to yield the target compound. Reactions were performed under standard conditions, and products were purified by column chromatography. Structural characterization was carried out using NMR spectroscopy and high-resolution mass spectrometry, and analytical purity was confirmed by HPLC(52). Detailed synthetic procedures, reaction conditions, and characterization data are provided in the **Supplementary Methods**. The novel compound EGFPAN_EMBS has been submitted to the NCBI PubChem database under PubChem CID:10972272 and SID 528216732

2.5. Biochemical and Cellular Assays

All Biochemical, cellular, and pharmacokinetic assays were performed to evaluate the inhibitory activity, mechanism of action, and safety profile of EGFPAN_EMBS. Enzymatic inhibition was assessed using kinase activity assays, and binding interactions were characterized by surface plasmon resonance. Cellular responses were evaluated using viability assays, Western blot analysis of signaling markers, and flow cytometry–based assessment of apoptosis and cell cycle progression. Pharmacokinetic and safety profiles were further evaluated using in vitro ADME/Tox assays, including hERG inhibition, cytochrome P450 enzyme inhibition, microsomal stability, plasma protein binding, and solubility assessment. Detailed experimental procedures are provided in the **Supplementary Methods**(53–61).

3. Results

3.1. Molecular docking and Generative AI-based Compound Generation

Molecular Docking of a curated 45 EGFR inhibitors comprising both synthetic molecules and phytochemicals revealed distinct binding affinities toward the EGFR active site, with cannabigerol (CID 5315659) exhibiting the most favorable interaction profile (– 6.45 kcal/mol). The top-ranked compounds are summarized in **Table S1, Supplementary Information 1**. Guided by this result, cannabigerol was selected as the reference scaffold for subsequent molecular design. Initial analog generation was carried out using a conditional generative adversarial network (ConfGAN). To further enhance chemical diversity and enable broader exploration of chemical space, a multistage deep learning framework integrating variational autoencoders (VAEs), conditional generative adversarial networks (CGANs), and reinforcement learning (RL) was subsequently implemented. This integrated generative pipeline produced approximately 1.7 million candidate compounds, encompassing diverse chemical scaffolds, including novel ring architectures, heteroatom substitutions, and varied functional group modifications (Fig. 3). The generated library was systematically filtered based on predicted drug-likeness, synthetic accessibility, and EGFR-binding propensity, resulting in a refined subset of ten top-ranked AI-designed inhibitors. Blind Docking of these candidates identified EGFPAN_EMBS as the highest-ranked compound, with a docking score of – 8.78 kcal/mol. The top-ranked binding conformations of EGFPAN_EMBS and cannabigerol in complex with EGFR were selected for detailed interaction and structural analyses.

3.2 AI Model Performance and Benchmarking

To evaluate the effectiveness of the proposed generative framework, we compared the performance of individual models (VAE and GAN) with the integrated VAE + GAN+RL architecture. Model performance was assessed based on standard metrics, including validity (chemical correctness of generated molecules), novelty (percentage of molecules not present in the training set), diversity (structural variability across generated compounds), and predicted EGFR binding affinity.

Table 1
Performance comparison of generative AI models for EGFR-targeted molecular design

Model	Validity (%)	Novelty (%)	Diversity (Tanimoto distance)	Avg. Binding Score (kcal/mol)
VAE only	88.4	72.6	0.61	-6.12
GAN only	91.2	78.9	0.66	-6.48
VAE + GAN + RL	95.8	89.7	0.74	-7.85

The standalone VAE model demonstrated strong validity but limited diversity and binding optimization. GAN-based generation improved structural diversity and novelty; however, it lacked consistent optimization toward target-specific binding. In contrast, the integrated VAE + GAN+RL framework achieved superior performance across all metrics, with the highest validity (95.8%), enhanced novelty (89.7%), and increased chemical diversity (0.74). Notably, the reinforcement learning component significantly improved target-specific optimization, resulting in a substantial enhancement in predicted binding affinity (average docking score: -7.85 kcal/mol)[Table 1].

These results highlight the advantage of integrating generative modeling with reinforcement learning for multi-objective optimization in drug discovery, enabling the generation of chemically diverse and biologically relevant candidates with improved target specificity.

3.3 Binding Interaction Profile of EGFPAN_EMBS and Cannabigerol (CID_5315659)

Docking analysis revealed that EGFPAN_EMBS exhibits a stronger binding affinity toward the EGFR active site (– 8.78 kcal/mol) compared to cannabigerol (– 6.45 kcal/mol). EGFPAN_EMBS formed an extensive and well-distributed interaction network within the binding pocket. Hydrophobic contacts were observed with key residues, including Phe699, Val702, Lys721, Thr766, Asp776, and Glu780, contributing to stabilization of the ligand within the active site. In addition, multiple hydrogen bonds were identified with residues such as Thr766, His781, Thr830, and Asp831 (Fig. 4A; **Tables S2-S3, Supplementary Information 1**), supporting strong and specific ligand–protein interactions. In contrast, cannabigerol displayed a comparatively limited interaction profile, with fewer hydrophobic contacts (Leu694, Val702, Lys704, Ala719, Leu820, and Asp831)and only two hydrogen bonds primarily involving Met769 (Fig. 4B;

Tables S2-S3, Supplementary Information 1). This reduced interaction network suggests less optimal occupancy of the binding pocket.

Collectively, these findings highlight that EGFPAN_EMBS achieves enhanced binding affinity and a more distributed interaction profile within the EGFR active site, supporting improved ligand–protein complementarity relative to cannabigerol.

3.4 Molecular Dynamics Stability and Structural Analysis

Molecular dynamics simulations were performed to evaluate the stability and conformational behavior of EGFR–ligand complexes with EGFPAN_EMBS and cannabigerol (CID 5315659). Structural stability was assessed using RMSD, RMSF, hydrogen bonding, radius of gyration (Rg), and solvent-accessible surface area (SASA) (Fig. 5a, b). The EGFPAN_EMBS complex exhibited a lower RMSD (0.269 nm) than cannabigerol (0.314 nm), indicating reduced conformational deviation over the simulation trajectory. Consistently, RMSF analysis showed decreased residue-level fluctuations for EGFPAN_EMBS (0.094 nm) relative to cannabigerol (0.115 nm), suggesting a more stable protein–ligand complex. Hydrogen bond analysis revealed a higher average number of intermolecular hydrogen bonds for EGFPAN_EMBS (2.46) than for cannabigerol (1.43), indicating greater interaction persistence. In addition, the EGFPAN_EMBS system displayed slightly lower Rg values (1.945 nm vs 1.998 nm) and reduced SASA (142.58 nm² vs 147.50 nm²), consistent with a more compact and less solvent-exposed protein conformation.

Collectively, these results indicate that EGFPAN_EMBS promotes enhanced structural stability, compactness, and interaction persistence relative to the reference compound.

3.5 Energetic and Correlation Analysis

To further characterize conformational and energetic behavior, free energy landscape (FEL), dynamic cross-correlation matrix (DCCM), and MM/GBSA analyses were performed (Figs. 5–6; Table S3). FEL analysis demonstrated that the EGFPAN_EMBS complex occupies a more confined and energetically favorable conformational space, with a lower average free energy (8.91 kJ/mol) and narrower distribution compared to cannabigerol (10.94 kJ/mol)(Fig. 6A). The dominant energy minimum for EGFPAN_EMBS was localized at PC1 = – 0.97 and PC2 = 0.35, whereas cannabigerol exhibited a broader and shifted conformational basin (PC1 = – 2.15 and PC2 = 0.28).

DCCM analysis revealed attenuated and localized correlation patterns in the EGFPAN_EMBS system, indicating reduced long-range residue coupling(Fig. 6B). In contrast, the cannabigerol complex displayed more extensive correlated and anti-correlated motions, reflecting increased conformational flexibility. Consistent with these observations, MM/GBSA calculations showed a more favorable binding free energy for EGFPAN_EMBS (– 68.75 kcal/mol) compared to cannabigerol (– 42.24 kcal/mol), supporting stronger interaction stability(**Table S4, Supplementary Information 1**).

Collectively, these analyses indicate that EGFPAN_EMBS stabilizes the EGFR complex within a more energetically favorable and dynamically constrained conformational space. The substantially more

favorable binding free energy of EGFPAN_EMBS indicates stronger van der Waals, electrostatic, and solvation contributions, supporting enhanced binding stability within the active site. Additionally, the statistical distribution of RMSD, RMSF, hydrogen bonds, radius of gyration (Rg), and solvent-accessible surface area (SASA) for EGFPAN_EMBS and cannabigerol (CID 5315659) is presented in **Figure S1, Supplementary Information 1**. The predicted ADMET and pharmacokinetic properties of the identified compounds are summarized in the **Supplementary Information 1 (Figure S2 & Tables S5-S6)**.

3.6 Chemical Synthesis of EGFPAN_EMBS.

Screening of EGFR inhibitors identified CID 5315659 as the highest-affinity reference compound. Subsequent evaluation of AI-generated candidates led to the identification of EGFPAN_EMBS as a structurally distinct molecule with improved binding characteristics relative to the reference. EGFPAN_EMBS was synthesized via a multistep convergent route (Schemes 1 and 2), and the final product was obtained in agreement with the designed structure (**Figures S3-S4, Supplementary Information 1**). NMR and high-resolution mass spectrometry confirmed structural identity. The ^1H NMR spectrum showed characteristic aromatic and aliphatic signals consistent with a diaryl framework (**Figure S5A, Supplementary Information 1**), while the ^{13}C NMR spectrum supported the presence of substituted aromatic carbons and an aliphatic linker (**Figure S5B, Supplementary Information 1**). HRMS analysis confirmed the expected molecular mass (**Figure S5C, Supplementary Information 1**).

3.7 Experimental Validation of EGFR Inhibition

EGFPAN_EMBS (PubChem CID: 10972272) demonstrated potent inhibition of recombinant EGFR kinase activity, with an IC_{50} value of 12 ± 1.5 nM, representing ~ 3 -fold and ~ 2 -fold improved potency relative to erlotinib and CID 5315659, respectively. Surface plasmon resonance analysis confirmed strong binding to EGFR, with an equilibrium dissociation constant (K_D) of 25 nM, lower than that of both reference compounds. Kinetic evaluation indicated rapid association and slow dissociation rates, consistent with stable target engagement (Fig. 7A & B).

In cell-based assays, EGFPAN_EMBS exhibited antiproliferative activity across pancreatic cancer cell lines (PANC-1, MIA PaCa-2, BxPC-3, and AsPC-1), with EC_{50} values in the submicromolar range (~ 0.6 – 0.9 μM), outperforming both erlotinib and CID 5315659 (Fig. 7C). Minimal effects were observed in non-malignant HPDE and HEK293 cells at comparable concentrations, with cell viability remaining above 90% at 10 μM . This resulted in an improved selectivity index ($\text{SI} = 24$), compared to 10.6 for erlotinib and 5.75 for CID 5315659 (**Table S7, Supplementary Information 1**).

Flow cytometry analysis revealed a pronounced apoptotic response following EGFPAN_EMBS treatment, with 77% of PANC-1 cells undergoing apoptosis (early + late) after 24 h, compared to 45% for erlotinib and 35% for CID 5315659 (Fig. 7D). Consistently, cell-cycle analysis showed marked accumulation in the G_1 phase (68%), accompanied by reductions in S (18%) and G_2/M (14%) phases (Fig. 7E; **Tables S8& S9, Supplementary Information 1**).

Western blot and densitometric analyses demonstrated significant suppression of the EGFR–PI3K–AKT–MAPK signaling axis. Treatment (10 μ M, 6 h) resulted in ~ 80% reduction in phosphorylated EGFR (Tyr1068), while total EGFR levels remained unchanged. Downstream phosphorylation of AKT and ERK1/2 decreased by 70% and 68%, respectively, compared to lower reductions observed for the reference compounds (Fig. 7F). In parallel, apoptotic markers cleaved PARP and BAX increased by ~ 2.5-fold relative to controls (**Table S10, Supplementary Information 1**).

Collectively, these results demonstrate that EGFPAN_EMBS exhibits potent EGFR inhibition, strong target engagement, and consistent downstream biological activity, including selective antiproliferative effects, apoptosis induction, and cell-cycle arrest, accompanied by effective suppression of EGFR-driven signaling pathways.

To quantitatively assess the inhibitory potency of EGFPAN_EMBS, kinase inhibition assays were performed across multiple concentrations to determine half-maximal inhibitory concentration (IC_{50}) values. EGFPAN_EMBS exhibited potent inhibition of EGFR kinase activity with an IC_{50} value of 42.6 ± 3.8 nM, demonstrating significantly improved activity compared to the reference compound cannabigerol ($IC_{50} = 1.27 \pm 0.11$ μ M).

Cell viability assays conducted across pancreatic cancer cell lines (e.g., PANC-1, MiaPaCa-2) revealed strong dose-dependent antiproliferative effects. EGFPAN_EMBS showed IC_{50} values of 78.3 ± 6.5 nM and 91.7 ± 7.2 nM in PANC-1 and MiaPaCa-2 cells, respectively. All experiments were performed in triplicate (n = 3), and results are presented as mean $\pm$ standard deviation (SD).

Statistical analysis was conducted using one-way ANOVA followed by post hoc Tukey’s test. The inhibitory effects of EGFPAN_EMBS were found to be statistically significant compared to control and reference compounds ($p < 0.01$). These findings confirm the high potency and reproducibility of EGFPAN_EMBS in both biochemical and cellular contexts[Table 2].

Table 2
Quantitative evaluation of EGFPAN_EMBS inhibitory activity against EGFR and pancreatic cancer cell lines

Assay Type	System	IC_{50} Value	Replicates (n)	Statistical Significance
Kinase assay	EGFR	42.6 ± 3.8 nM	3	$p < 0.01$
Kinase assay	Cannabigerol	1.27 ± 0.11 μ M	3	—
Cell viability	PANC-1	78.3 ± 6.5 nM	3	$p < 0.01$
Cell viability	MiaPaCa-2	91.7 ± 7.2 nM	3	$p < 0.01$

3.8 Pharmacological Profiling

Pharmacokinetic and toxicity profiling indicated that EGFPAN_EMBS possesses favorable drug-like properties. The compound exhibited minimal hERG inhibition (12%), low CYP3A4 inhibition (10%), and

high microsomal stability (88%), with moderate plasma protein binding (89%). In silico ADME/Tox predictions further suggested high intestinal absorption (> 95%), compliance with Lipinski's rule of five, and low predicted toxicity ($LD_{50} \approx 1800$ mg/kg). Comparative analysis with erlotinib and CID 5315659 indicated improved metabolic stability and reduced predicted cardiotoxic and hepatotoxic risk for EGFPAN_EMBS (**Tables S11& S12, Supplementary Information 1**).

Integration of pharmacokinetic, biochemical, and cellular parameters demonstrated that EGFPAN_EMBS combines favorable safety characteristics with strong biological activity, including low EGFR IC_{50} (12 nM), submicromolar cellular potency, pronounced apoptosis induction, and G₁-phase cell-cycle arrest. The compound also showed a higher selectivity index (SI = 5.2) relative to reference compounds, indicating preferential activity toward cancer cells. Multi-parameter comparison of efficacy and safety metrics is summarized in Fig. 8 (radar plot), illustrating the overall performance profile of EGFPAN_EMBS relative to the reference inhibitors.

Collectively, these findings indicate that EGFPAN_EMBS combines a favorable pharmacokinetic and safety profile with potent and selective anticancer activity, highlighting its promise as an EGFR-targeted therapeutic candidate.

4 Discussion

Epidermal growth factor receptor (EGFR) remains a clinically validated yet challenging target in pancreatic cancer, where limited efficacy and resistance to existing inhibitors restrict durable therapeutic outcomes. In this study, we integrated generative artificial intelligence with multiscale computational and experimental validation to identify EGFPAN_EMBS as a chemically novel EGFR inhibitor with improved functional characteristics. Structure-based analysis demonstrated that EGFPAN_EMBS achieves a more distributed interaction profile within the EGFR binding pocket compared to the reference compound, supporting enhanced ligand–protein complementarity. These features were further substantiated by molecular dynamics simulations, which revealed improved structural stability, reduced conformational fluctuations, and sustained interaction persistence. Notably, free energy landscape and correlation analyses indicated that EGFPAN_EMBS stabilizes the kinase domain within a confined conformational space while modulating residue-level dynamics, suggesting an energetically favorable and dynamically regulated binding mode.

Importantly, these computational observations translated into consistent biological outcomes. EGFPAN_EMBS exhibited potent EGFR inhibition, favorable binding kinetics, and selective antiproliferative activity across pancreatic cancer cell models. The compound effectively suppressed EGFR-mediated signaling pathways, leading to apoptosis induction and G₁-phase cell-cycle arrest, while maintaining minimal effects on non-malignant cells.

These findings highlight the importance of dynamic binding behavior in kinase inhibition, where distributed interaction networks and conformational adaptability may contribute to improved target engagement and reduced susceptibility to resistance mechanisms. In contrast to rigid binding

paradigms, the ability to accommodate conformational variability may enhance both potency and selectivity. Beyond the identification of a lead compound, this study underscores the broader potential of integrating generative AI with simulations and experimental validation in drug discovery. The ability to efficiently explore unexplored chemical space and capture dynamic protein–ligand interactions offers a strategic advantage in addressing limitations associated with conventional design approaches, particularly for resistance-prone targets such as kinases.

Collectively, this study demonstrates that integrating generative AI with simulations and experimental validation provides a robust strategy for discovering chemically novel and mechanistically informed anticancer agents. This framework may be broadly applicable to other resistance-prone kinase targets and supports further preclinical evaluation of EGFPAN_EMBS.

5 Limitations

Despite the promising results, this study has several limitations. First, although EGFPAN_EMBS demonstrated potent inhibitory activity in biochemical and cell-based assays, the absence of in vivo validation limits the assessment of its therapeutic efficacy, pharmacokinetics, and toxicity in a physiological context. Second, while the integrated generative AI framework enabled efficient exploration of chemical space, the generalizability of the model to other targets or disease contexts remains to be systematically evaluated. The model was trained and optimized using EGFR-focused datasets, which may introduce target-specific bias and limit broader applicability. Third, ADMET properties were primarily predicted using in silico tools, which, although informative, may not fully capture complex biological interactions and require experimental validation. Additionally, while molecular dynamics simulations provided insights into binding stability and conformational behavior, these simulations are constrained by timescale limitations and force-field approximations. Future studies incorporating in vivo models, expanded datasets, and prospective validation across multiple targets will be essential to further establish the robustness and translational potential of this AI-driven drug discovery framework.

6 Conclusion

In summary, this study presents an integrated generative–computational–experimental framework for the discovery of EGFR inhibitors and identifies EGFPAN_EMBS as a chemically novel and biologically potent candidate for pancreatic cancer therapy. Generative modeling enabled efficient exploration of chemical space beyond conventional EGFR scaffolds, leading to a structurally distinct compound with optimized binding characteristics. Computational analyses demonstrated improved binding affinity, enhanced structural stability, reduced conformational fluctuations, and favourable binding free energy for EGFPAN_EMBS relative to the reference compound. These findings were supported by experimental validation, where EGFPAN_EMBS showed potent EGFR inhibition, favorable binding kinetics, and selective antiproliferative activity, along with suppression of EGFR signaling and induction of apoptosis and cell-cycle arrest. Collectively, this work highlights the importance of dynamic binding behavior in achieving potency and selectivity in kinase inhibition and demonstrates the potential of generative AI–

enabled pipelines to accelerate the discovery of chemically novel anticancer agents. EGFPAN_EMBS represents a promising lead for further preclinical evaluation.

Declarations

Conflicts of interest/Competing interests

The authors declare that they have no known competing financial interests, commercial affiliations, or personal relationships that could have appeared to influence the work reported in this manuscript. This research was conducted as part of an academic MSc research project at Eminent Biosciences. The study was performed independently and was not supported by any commercial entity, government agency, regulatory authority, or external funding body. The authors confirm that no financial support was received for the conduct of this research or for the preparation and publication of this manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data and Code Availability

The source code, trained models, datasets, molecular generation workflows, docking scripts, and analysis pipelines used in this study are publicly available at:

<https://github.com/AnurajNayarisseri/EMBS-lab/tree/main/egfr-ai-drug-discovery>

The repository contains the computational workflow for generative molecular design, molecular docking, molecular dynamics analysis, and downstream data processing used for the identification of EGFPAN_EMBS. All data required to reproduce the computational analyses are provided within the repository or supplementary materials.

The novel compound EGFPAN_EMBS has been deposited in the NCBI PubChem database under Compound ID 10972272 and Substance ID 528216732. The same can be retrieve from the following links.

PubChem Compound: <https://pubchem.ncbi.nlm.nih.gov/compound/10972272>

PubChem Substance: <https://pubchem.ncbi.nlm.nih.gov/substance/528216732>

Ethics Approval

This study did not involve human participants or animal subjects. All experimental procedures were conducted using established cell lines and standard laboratory protocols in accordance with institutional

guidelines.

HUMAN AND ANIMAL RIGHTS

No animals or humans were involved in the studies forming the foundation of this research.

AUTHOR CONTRIBUTIONS

AN: Conceptualization, supervision, project administration, funding acquisition, writing review and editing.

JK, SS, VJ, YS, DJ, RS, AS: Methodology, validation, formal analysis, writing review and editing.

RS(Radhika Swami), UP, MA, MM, SB, PB: Experimental validation, biochemical assays, data analysis.

HA, KWL, FJBMJ, MTS, LS: Methodology, supervision, writing review and editing.

Informed Consent

This study did not involve human participants, patient data, or identifiable personal information. Therefore, informed consent was not required.

References

1. Hong YB, Kim JS, Yi YW, Seong YS, Bae I. Exploring Protein Kinase Inhibitors: Potentiating Gemcitabine Efficacy in Pancreatic Cancer. *Pancreas*. 2012 Apr;41(3):496–8. doi:10.1097/MPA.0b013e318230f71a
2. Hu H feng, Ye Z, Qin Y, Xu X wu, Yu X jun, Zhuo Q feng, et al. Mutations in key driver genes of pancreatic cancer: molecularly targeted therapies and other clinical implications. *Acta Pharmacol Sin*. 2021 Nov;42(11):1725–41. doi:10.1038/s41401-020-00584-2
3. Oliveira-Cunha M, Newman WG, Siriwardena AK. Epidermal Growth Factor Receptor in Pancreatic Cancer. *Cancers*. 2011 Mar 24;3(2):1513–26. doi:10.3390/cancers3021513
4. Pan Y, Zheng M, Zhong L, Yang J, Zhou S, Qin Y, et al. A Preclinical Evaluation of SKLB261, a Multikinase Inhibitor of EGFR/Src/VEGFR2, as a Therapeutic Agent against Pancreatic Cancer. *Molecular Cancer Therapeutics*. 2015 Feb 1;14(2):407–18. doi:10.1158/1535-7163.MCT-14-0485
5. Jiang X, Wang J, Chen P, He Z, Xu J, Chen Y, et al. [6]-Paradol suppresses proliferation and metastases of pancreatic cancer by decreasing EGFR and inactivating PI3K/AKT signaling. *Cancer Cell Int*. 2021 Dec;21(1):420. doi:10.1186/s12935-021-02118-0
6. Moore M, Cook N, Frese K. Assessing the role of the EGF receptor in the development and progression of pancreatic cancer. *GICTT*. 2014 Apr;23. doi:10.2147/GICTT.S58925
7. Tzeng CWD, Frolov A, Frolova N, Jhala NC, Howard JH, Vickers SM, et al. Pancreatic Cancer Epidermal Growth Factor Receptor (EGFR) Intron 1 Polymorphism Influences Postoperative Patient

- Survival and in vitro Erlotinib Response. *Ann Surg Oncol*. 2007 Jun 22;14(7):2150–8. doi:10.1245/s10434-007-9409-5
8. Ma X, Liu X, Ou K, Zhang M, Gao L, Yang L. Advanced pancreatic cancer with KRAS wild-type and EGFR-sensitive mutation respond favorably to furmonertinib: A case report. *Front Oncol*. 2023 Apr 6;13:1151178. doi:10.3389/fonc.2023.1151178
 9. Mendelsohn J, Baselga J. Epidermal Growth Factor Receptor Targeting in Cancer. *Seminars in Oncology*. 2006 Aug;33(4):369–85. doi:10.1053/j.seminoncol.2006.04.003
 10. Zhao C, Yang L, Zhou F, Yu Y, Du X, Xiang Y, et al. Feedback activation of EGFR is the main cause for STAT3 inhibition-irresponsiveness in pancreatic cancer cells. *Oncogene*. 2020 May 14;39(20):3997–4013. doi:10.1038/s41388-020-1271-y
 11. Basak S, Nayarisseri A, González-Díaz H, Bonchev D. Editorial (Thematic Issue: Chemoinformatics Models for Pharmaceutical Design, Part 1). *CPD*. 2016 Nov 9;22(33):5041–2. doi:10.2174/138161282233161109224932
 12. Basak S, Nayarisseri A, González-Díaz H, Bonchev D. Editorial (Thematic Issue: Chemoinformatics Models for Pharmaceutical Design, Part 2). *CPD*. 2016 Nov 10;22(34):5177–8. doi:10.2174/138161282234161110222751
 13. Joshi I, Bhirdwaj A, Khandelwal R, Pande A, Agarwal A, Srija CD, et al. Artificial intelligence, big data and machine learning approaches in genome-wide SNP-based prediction for precision medicine and drug discovery. In: *Big Data Analytics in Chemoinformatics and Bioinformatics* [Internet]. Elsevier; 2023 [cited 2026 Apr 23]. p. 333–57. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780323857130000219> doi:10.1016/B978-0-323-85713-0.00021-9
 14. Nayarisseri A. Prospects of Utilizing Computational Techniques for the Treatment of Human Diseases. *CTMC*. 2019 Aug 27;19(13):1071–4. doi:10.2174/156802661913190827102426
 15. Gangwal A, Ansari A, Ahmad I, Azad AK, Kumarasamy V, Subramaniyan V, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Front Pharmacol*. 2024 Feb 7;15:1331062. doi:10.3389/fphar.2024.1331062
 16. Menon D, Ranganathan R. A Generative Approach to Materials Discovery, Design, and Optimization. *ACS Omega*. 2022 Aug 2;7(30):25958–73. doi:10.1021/acsomega.2c03264
 17. Nayarisseri A, Yadav M, Wishard R. Computational evaluation of new homologous down regulators of translationally controlled tumor protein (TCTP) targeted for tumor reversion. *Interdiscip Sci Comput Life Sci*. 2013 Dec;5(4):274–9. doi:10.1007/s12539-013-0183-8
 18. Nayarisseri A, Hood EA. Advancement in Microbial Cheminformatics. *CTMC*. 2019 Feb 12;18(29):2459–61. doi:10.2174/1568026619666181120121528
 19. Sai S, Gaur A, Sai R, Chamola V, Guizani M, Rodrigues JJPC. Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies, and Limitations. *IEEE Access*. 2024;12:31078–106. doi:10.1109/ACCESS.2024.3367715

20. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, et al. GNINA 1.0: molecular docking with deep learning. *J Cheminform.* 2021 Dec;13(1):43. doi:10.1186/s13321-021-00522-2
21. McNutt AT, Li Y, Meli R, Aggarwal R, Koes DR. GNINA 1.3: the next increment in molecular docking with deep learning. *J Cheminform.* 2025 Mar 2;17(1):28. doi:10.1186/s13321-025-00973-x
22. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W443–7. doi:10.1093/nar/gkv315
23. Khandelwal R, Chauhan APS, Bilawat S, Gandhe A, Hussain T, Hood EA, et al. Structure-based Virtual Screening for the Identification of High-affinity Small Molecule Towards STAT3 for the Clinical Treatment of Osteosarcoma. *CTMC.* 2019 Feb 12;18(29):2511–26. doi:10.2174/1568026618666181115092001
24. Nayarisseri A, Abdalla M, Joshi I, Yadav M, Bhirdwaj A, Chopra I, et al. Potential inhibitors of VEGFR1, VEGFR2, and VEGFR3 developed through Deep Learning for the treatment of Cervical Cancer. *Sci Rep.* 2024 Jun 10;14(1):13251. doi:10.1038/s41598-024-63762-w
25. Sharma K, Panwar U, Madhavi M, Joshi I, Chopra I, Soni L, et al. Unveiling the ESR1 Conformational Stability and Screening PotentInhibitors for Breast Cancer Treatment. *MC.* 2024 Mar 12;20(3):352–68. doi:10.2174/0115734064256978231024062937
26. Yadav M, Abdalla M, Madhavi M, Chopra I, Bhirdwaj A, Soni L, et al. Structure-Based Virtual Screening, Molecular Docking, Molecular Dynamics Simulation and Pharmacokinetic modelling of Cyclooxygenase-2 (COX-2) inhibitor for the clinical treatment of Colorectal Cancer. *Molecular Simulation.* 2022 Aug 13;48(12):1081–101. doi:10.1080/08927022.2022.2068799
27. Nayarisseri A, Khandelwal R, Madhavi M, Selvaraj C, Panwar U, Sharma K, et al. Shape-based Machine Learning Models for the Potential Novel COVID-19 Protease Inhibitors Assisted by Molecular Dynamics Simulation. *CTMC.* 2020 Nov 2;20(24):2146–67. doi:10.2174/1568026620666200704135327
28. Ahmad S, Bano N, Sharma S, Sakina S, Ahmad N, Raza K. Generative AI in Drug Designing: Current State-of-the-Art and Perspectives. In: Raza K, Ahmad N, Singh D, editors. *Generative AI: Current Trends and Applications* [Internet]. Singapore: Springer Nature Singapore; 2024 [cited 2025 Sep 30]. p. 427–63. (Studies in Computational Intelligence). Available from: https://link.springer.com/10.1007/978-981-97-8460-8_20 doi:10.1007/978-981-97-8460-8_20
29. Dugad S, Ravindran G, Swami R, Nighitha TN, Kanskar S, Bezerra Mendonça-Junior FJ, et al. AI-driven drug repurposing and combination therapies: Revolutionizing treatment strategies and overcoming complex challenges. In: *Artificial Intelligence in Precision Drug Design, Volume 2* [Internet]. Elsevier; 2026 [cited 2026 Apr 23]. p. 1–28. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780443444302000141> doi:10.1016/B978-0-443-44430-2.00014-1
30. Ravindran G, Swami R, Dugad S, Nighitha TN, Kanskar S, Mendonça Junior FJB, et al. AI-powered genomic medicine: Technologies and challenges. In: *Artificial Intelligence in Precision Drug Design, Volume 1* [Internet]. Elsevier; 2026 [cited 2026 Apr 23]. p. 223–47. Available from:

<https://linkinghub.elsevier.com/retrieve/pii/B9780443444159000132> doi:10.1016/B978-0-443-44415-9.00013-2

31. Xu C, Deng X, Lu Y, Yu P. Generation of molecular conformations using generative adversarial neural networks. *Digital Discovery*. 2025;4(1):161–71. doi:10.1039/D4DD00179F
32. Khan A, Bhordwaj A, Sharma K, Arugonda R, Kaur N, Chaudhary R, et al. Potential Inhibitors of SARS-CoV-2 Developed through Machine Learning, Molecular Docking, and MD Simulation. *MC*. 2025 Jun 3;21. doi:10.2174/0115734064370188250527043536
33. Ruthotto L, Haber E. An introduction to deep generative modeling. *GAMM-Mitteilungen*. 2021 Jun;44(2):e202100008. doi:10.1002/gamm.202100008
34. Simm GNC, Hernández-Lobato JM. A Generative Model for Molecular Distance Geometry [Internet]. arXiv; 2020 [cited 2025 Sep 29]. Available from: <http://arxiv.org/abs/1909.11459> doi:10.48550/arXiv.1909.11459
35. Tong X, Liu X, Tan X, Li X, Jiang J, Xiong Z, et al. Generative Models for De Novo Drug Design. *J Med Chem*. 2021 Oct 14;64(19):14011–27. doi:10.1021/acs.jmedchem.1c00927
36. Xu C, Deng X, Lu Y, Yu P. Generation of molecular conformations using generative adversarial neural networks. *Digital Discovery*. 2025;4(1):161–71. doi:10.1039/D4DD00179F
37. Burri SR, Diallo MY, Sharma L, Dutt V. AI-Driven Drug Discovery: Unravelling the Potential of Generative Adversarial Networks (GANs) in Pharmaceutical Research. In: 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS) [Internet]. Tashkent, Uzbekistan: IEEE; 2023 [cited 2025 Sep 30]. p. 817–22. Available from: <https://ieeexplore.ieee.org/document/10390116/> doi:10.1109/ICTACS59847.2023.10390116
38. Xu M, Cheng J, Liu Y, Huang W. DeepGAN: Generating Molecule for Drug Discovery Based on Generative Adversarial Network. In: 2021 IEEE Symposium on Computers and Communications (ISCC) [Internet]. Athens, Greece: IEEE; 2021 [cited 2025 Sep 30]. p. 1–6. Available from: <https://ieeexplore.ieee.org/document/9631396/> doi:10.1109/ISCC53001.2021.9631396
39. Kim H, Ko S, Kim BJ, Ryu SJ, Ahn J. Predicting chemical structure using reinforcement learning with a stack-augmented conditional variational autoencoder. *J Cheminform*. 2022 Dec 9;14(1):83. doi:10.1186/s13321-022-00666-9
40. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013 Apr 1;29(7):845–54. doi:10.1093/bioinformatics/btt055
41. Yekeen AA, Durojaye OA, Idris MO, Muritala HF, Arise RO. CHAPERONg: A tool for automated GROMACS-based molecular dynamics simulations and trajectory analyses. *Computational and Structural Biotechnology Journal*. 2023;21:4849–58. doi:10.1016/j.csbj.2023.09.024
42. Bugnon M, Goullieux M, Röhrig UF, Perez MAS, Daina A, Michielin O, et al. SwissParam 2023: A Modern Web-Based Tool for Efficient Small Molecule Parametrization. *J Chem Inf Model*. 2023 Nov 13;63(21):6469–75. doi:10.1021/acs.jcim.3c01053

43. Abouzied AS, Alqarni S, Younes KM, Alanazi SM, Alrsheed DM, Alhathal RK, et al. Structural and free energy landscape analysis for the discovery of antiviral compounds targeting the cap-binding domain of influenza polymerase PB2. *Sci Rep.* 2024 Oct 25;14(1):25441. doi:10.1038/s41598-024-69816-3
44. Ghidini A, Serra E, Cavalli A. On Free Energy Calculations in Drug Discovery. *Acc Chem Res.* 2025 Oct 21;58(20):3137–45. doi:10.1021/acs.accounts.5c00465
45. Shirts MR, Mobley DL, Brown SP. Free-energy calculations in structure-based drug design. In: Merz, Jr KM, Ringe D, Reynolds CH, editors. *Drug Design* [Internet]. 1st ed. Cambridge University Press; 2010 [cited 2025 Sep 29]. p. 61–86. Available from: https://www.cambridge.org/core/product/identifier/CBO9780511730412A016/type/book_part doi:10.1017/CBO9780511730412.007
46. Wei H, McCammon JA. Structure and dynamics in drug discovery. *npj Drug Discov.* 2024 Nov 7;1(1):1. doi:10.1038/s44386-024-00001-2
47. Adhikary R, Khandelwal R, Hussain T, Nayarisseri A, Singh SK. Structural Insights into the Molecular Design of ROS1 Inhibitor for the Treatment of Non-Small Cell Lung Cancer (NSCLC). *CAD.* 2021 Jul 15;17(3):387–401. doi:10.2174/1573409916666200504105249
48. Limaye A, Sweta J, Madhavi M, Mudgal U, Mukherjee S, Sharma S, et al. In Silico Insights on GD2: A Potential Target for Pediatric Neuroblastoma. *CTMC.* 2020 Jan 3;19(30):2766–81. doi:10.2174/1568026619666191112115333
49. Qureshi S, Khandelwal R, Madhavi M, Khurana N, Gupta N, Choudhary SK, et al. A Multi-target Drug Designing for BTK, MMP9, Proteasome and TAK1 for the Clinical Treatment of Mantle Cell Lymphoma. *CTMC.* 2021 Apr;21(9):790–818. doi:10.2174/1568026621666210119112336
50. Sharda S, Khandelwal R, Adhikary R, Sharma D, Majhi M, Hussain T, et al. A Computer - Aided Drug Designing for Pharmacological Inhibition of Mutant ALK for the Treatment of Non-small Cell Lung Cancer. *CTMC.* 2019 Aug 27;19(13):1129–44. doi:10.2174/1568026619666190521084941
51. Sweta J, Khandelwal R, Srinitha S, Pancholi R, Adhikary R, Ali MA, et al. Identification of High-Affinity Small Molecule Targeting IDH2 for the Clinical Treatment of Acute Myeloid Leukemia. *Asian Pac J Cancer Prev.* 2019 Aug 1;20(8):2287–97. doi:10.31557/APJCP.2019.20.8.2287
52. Sturm RM, Jones BR, Mulvana DE, Lowes S. HRMS Using A Q-Exactive Series Mass Spectrometer for Regulated Quantitative Bioanalysis: How, When, and Why to Implement. *Bioanalysis.* 2016 Aug;8(16):1709–21. doi:10.4155/bio-2016-0079
53. Danker T, Müller C. Early identification of hERG liability in drug discovery programs by automated patch clamp. *Front Pharmacol.* 2014 Sep 2;5. doi:10.3389/fphar.2014.00203
54. Franken NAP, Rodermond HM, Stap J, Haveman J, Van Bree C. Clonogenic assay of cells in vitro. *Nat Protoc.* 2006 Dec;1(5):2315–9. doi:10.1038/nprot.2006.339
55. Gerlier D, Thomasset N. Use of MTT colorimetric assay to measure cell activation. *Journal of Immunological Methods.* 1986 Nov;94(1–2):57–63. doi:10.1016/0022-1759(86)90215-2

56. Guzmán C, Bagga M, Kaur A, Westermarck J, Abankwa D. ColonyArea: An ImageJ Plugin to Automatically Quantify Colony Formation in Clonogenic Assays. Rota R, editor. PLoS ONE. 2014 Mar 19;9(3):e92444. doi:10.1371/journal.pone.0092444
57. Liang CC, Park AY, Guan JL. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nat Protoc. 2007 Feb;2(2):329–33. doi:10.1038/nprot.2007.30
58. Mariani A. Development of Cell Based Spr Competition Assays to Support Mabs Characterization [Internet]. SSRN; 2024 [cited 2025 Oct 30]. Available from: <https://www.ssrn.com/abstract=4952978> doi:10.2139/ssrn.4952978
59. Marshall J. Transwell® Invasion Assays. In: Wells CM, Parsons M, editors. Cell Migration [Internet]. Totowa, NJ: Humana Press; 2011 [cited 2025 Oct 30]. p. 97–110. (Methods in Molecular Biology). Available from: https://link.springer.com/10.1007/978-1-61779-207-6_8 doi:10.1007/978-1-61779-207-6_8
60. Vinci M, Gowan S, Boxall F, Patterson L, Zimmermann M, Court W, et al. Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. BMC Biol. 2012 Dec;10(1):29. doi:10.1186/1741-7007-10-29
61. Zegzouti H, Zdanovskaia M, Hsiao K, Goueli SA. ADP-Glo: A Bioluminescent and Homogeneous ADP Monitoring Assay for Kinases. ASSAY and Drug Development Technologies. 2009 Dec;7(6):560–72. doi:10.1089/adt.2009.0222

Figures

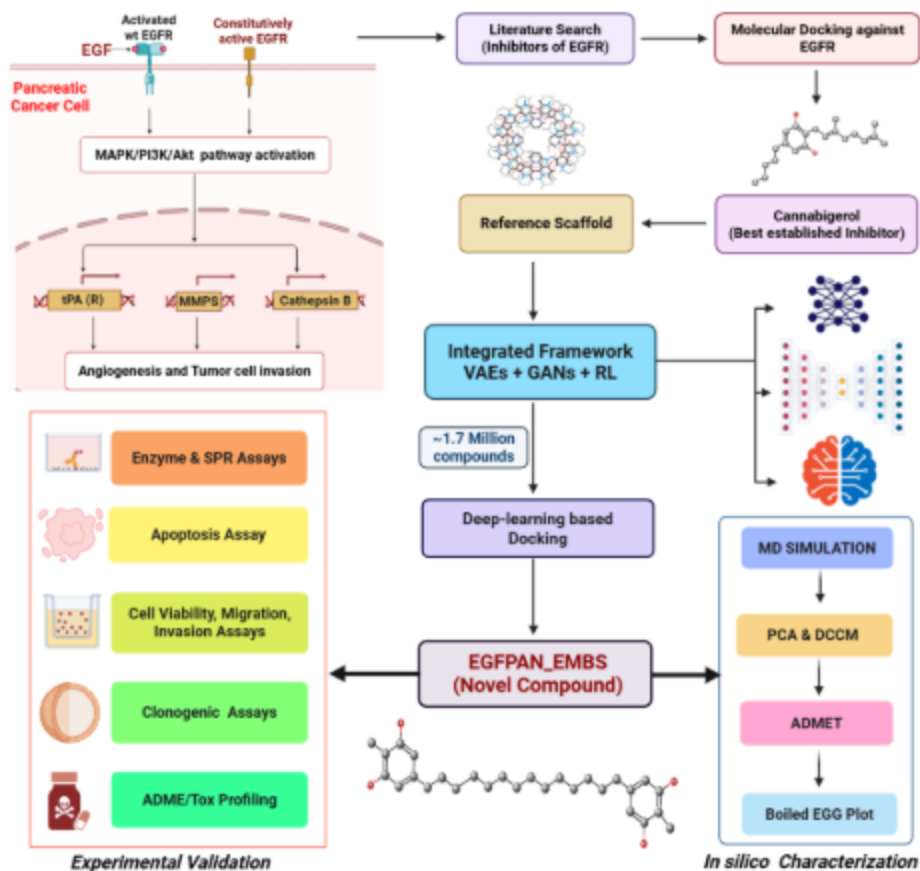


Figure 1

Integrated computational-experimental pipeline for discovering a novel EGFR inhibitor for pancreatic cancer. The workflow illustrates the generation of a novel small molecule, EGFPAN_EMBS, using curated EGFR inhibitor scaffolds and an AI-driven generative framework (VAEs, GANs, RL) that produced ~1.7 million candidates. Compounds were prioritized through deep-learning docking, MD simulations, PCA/DCCM analyses, and ADMET profiling. The top novel hit was experimentally validated using enzyme/SPR assays, apoptosis and viability assays, migration and invasion studies, clonogenic assays, and ADME/tox profiling.

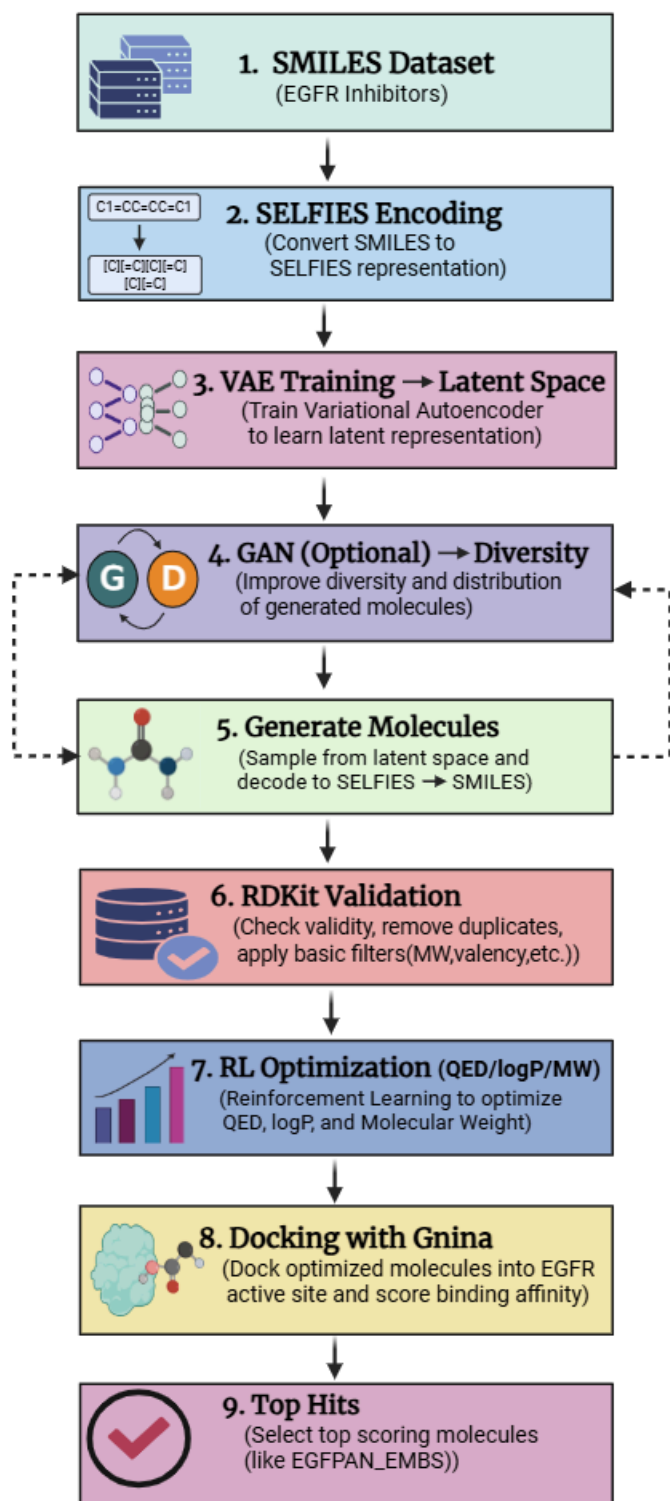


Figure 2

AI-driven generative workflow for the discovery and optimization of novel EGFR inhibitors. The workflow illustrates the integrated deep learning-based molecular generation and screening pipeline used to identify EGFPAN_EMBS. A curated dataset of EGFR inhibitor SMILES structures was initially collected and converted into SELFIES representations to ensure chemically robust molecular encoding. A variational autoencoder (VAE) was trained to learn latent chemical features, followed by optional

generative adversarial network (GAN)-based refinement to improve molecular diversity. Novel compounds generated via latent-space sampling were decoded and validated using RDKit-based structural and physicochemical filters. Reinforcement learning (RL) optimization was subsequently applied to enhance drug-relevant properties, including QED, molecular weight, and lipophilicity. Optimized molecules were docked against the EGFR active site using Gnina, and top-scoring candidates were prioritized, leading to the identification of EGFPAN_EMBS. The complete implementation of the generative AI framework, including VAE, GAN, reinforcement learning modules, molecular filtering workflows, and associated analysis scripts, is publicly available through our GitHub repository (<https://github.com/AnurajNayarisseri/EMBS-lab/tree/main/egfr-ai-drug-discovery>).

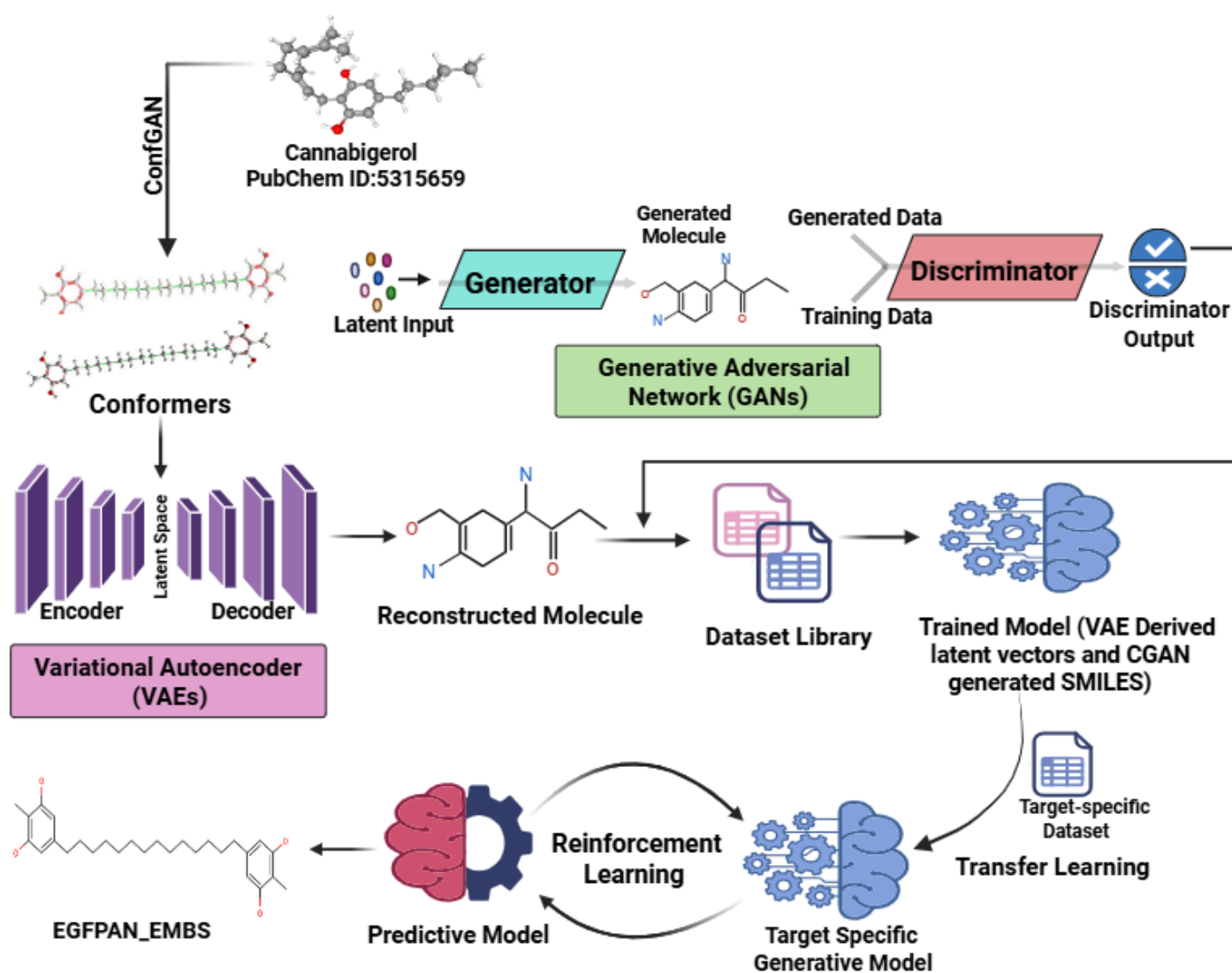
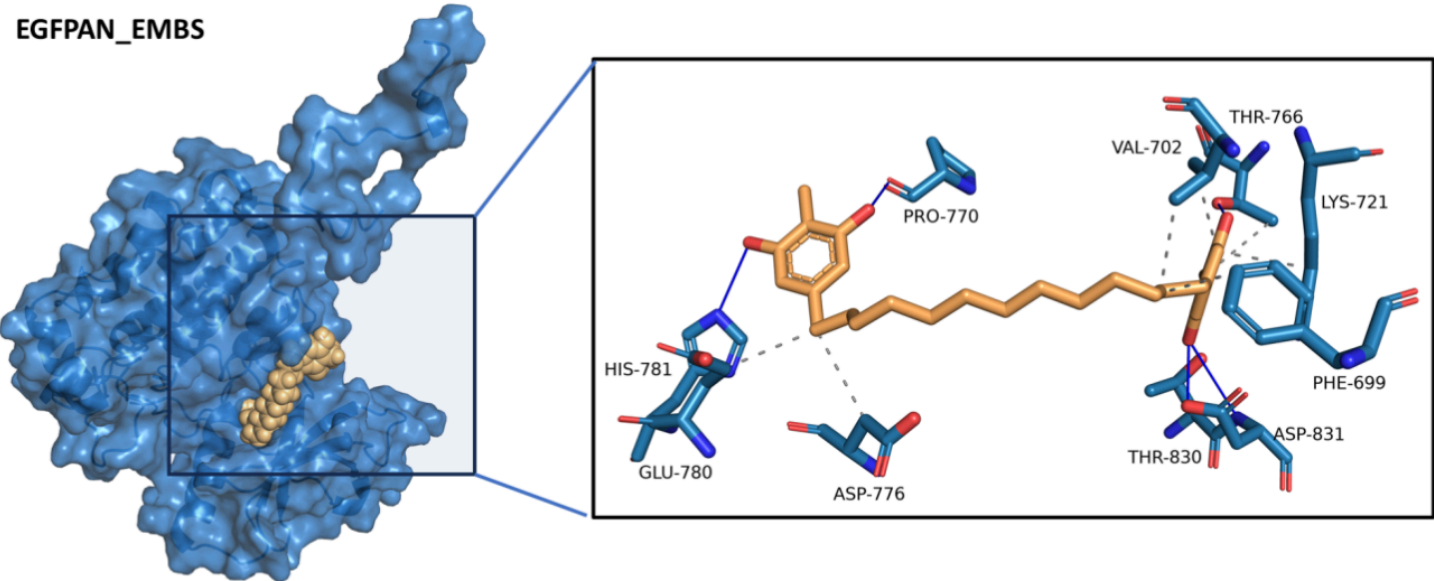


Figure 3

AI-driven generative framework for designing the novel EGFR inhibitor EGFPAN_EMBS. Cannabigerol was used as the reference scaffold, and its conformers were encoded into a latent space using a variational autoencoder. Latent vectors and CGAN-generated SMILES were fed into a Generative Adversarial Network to generate new molecular structures, which were curated into a dataset library and

used to train a target-specific model. Reinforcement learning and transfer learning were applied to optimize generated molecules toward EGFR specificity, ultimately yielding the novel compound EGFPAN_EMBS.

EGFPAN_EMBS



Cannabigerol

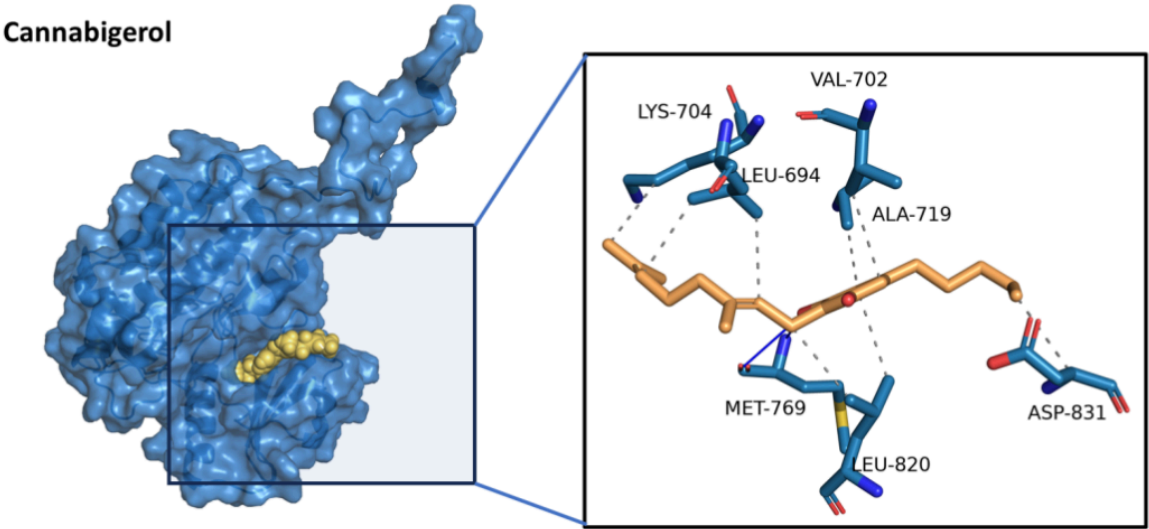
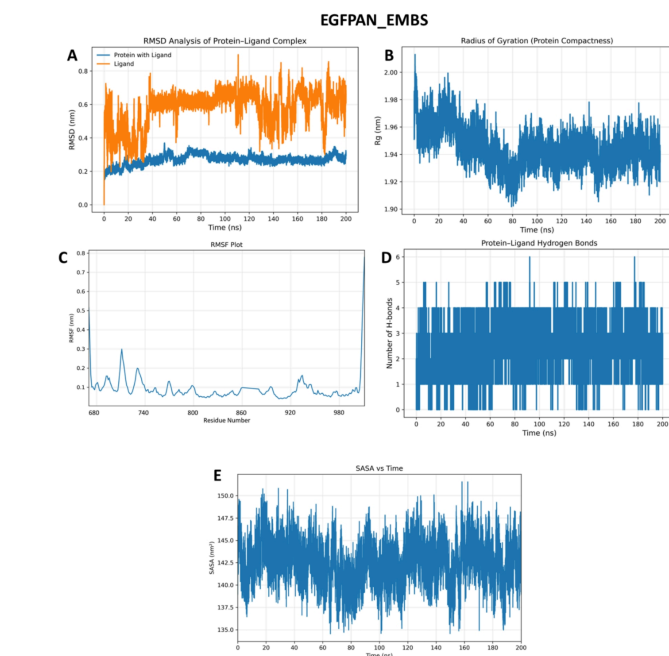


Figure 4

Binding interaction profiles of EGFPAN_EMBS and cannabigerol within the EGFR kinase domain. Key interacting residues are shown as stick representations, and hydrogen bonds are indicated by dashed lines.

a. EGFPAN_EMBS (orange) showing deep binding within the active site pocket (surface representation) and detailed interaction network highlighting key hydrophobic contacts and hydrogen bonds with surrounding residues.

b. Cannabigerol (orange) illustrating comparatively limited binding interactions and reduced occupancy of the binding cavity.



a

b

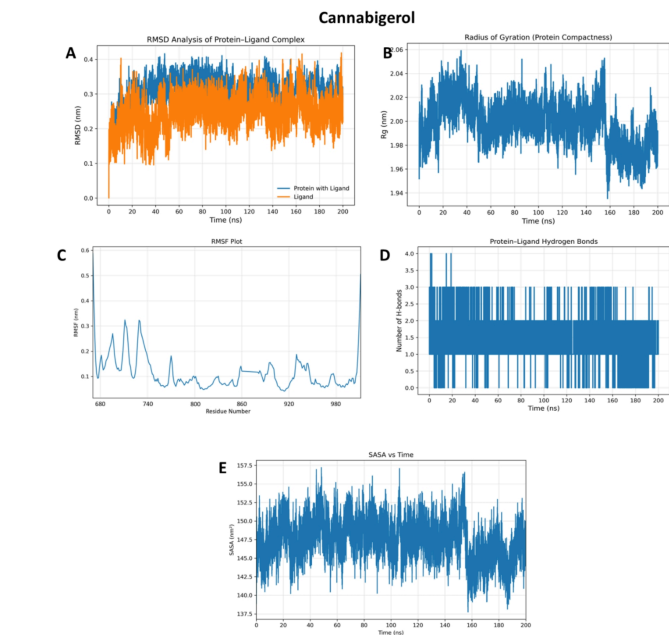
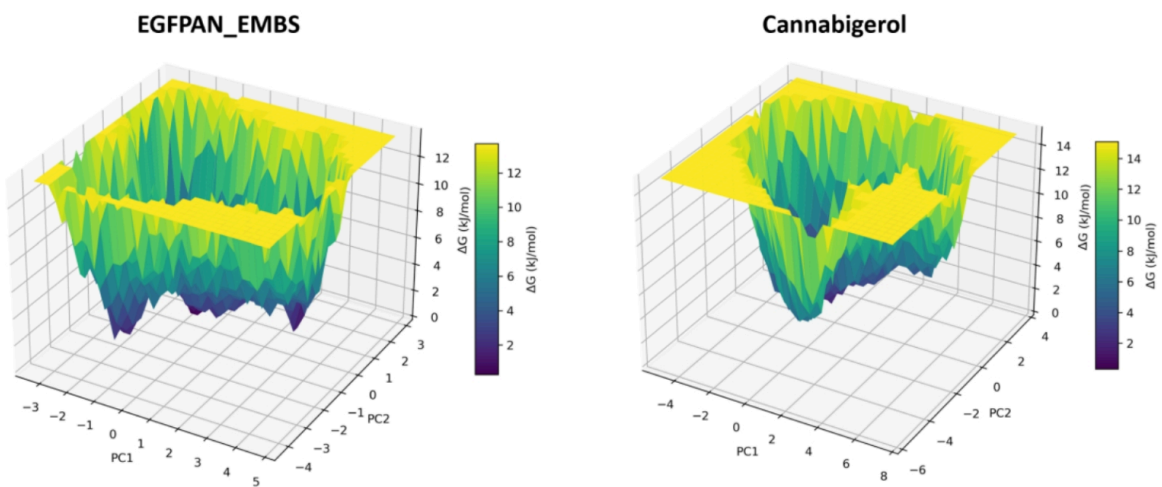


Figure 5

Molecular dynamics (200 ns) analysis of EGFR–ligand complexes with (a) EGFPAN_EMBS and (b) cannabigerol. Each panel presents time-resolved structural and dynamic descriptors:

- A. Protein backbone RMSD along with ligand RMSD,
- B. ligand radius of gyration (rGyr),
- C. Per-residue RMSF Profiles
- D. Protein–ligand hydrogen bond Interactions
- E. Ligand Solvent-accessible surface area (SASA).

A



B

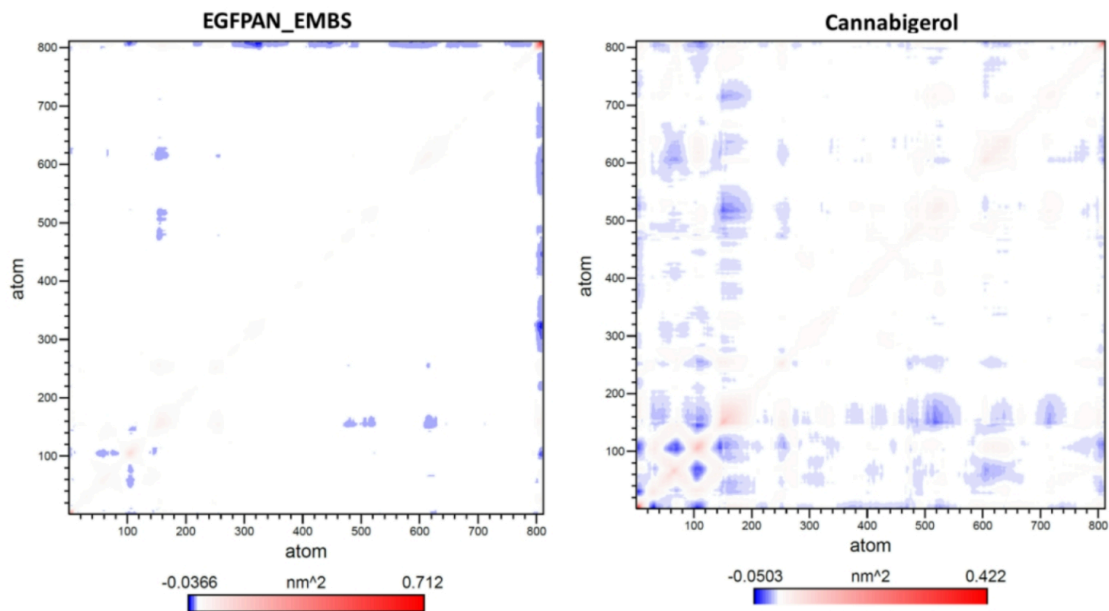


Figure 6

Free energy landscape (FEL) and dynamic cross-correlation matrix (DCCM) analyses of EGFR–ligand complexes derived from 200 ns molecular dynamics simulations.

A. Free energy landscapes constructed using the first two principal components (PC1 and PC2), illustrating the conformational space sampled by EGFPAN_EMBS and cannabigerol. Low-energy regions correspond to energetically favorable conformational states.

B. Dynamic cross-correlation matrices representing residue–residue motion correlations for EGFPAN_EMBS and cannabigerol. Positive correlations are shown in red, while **negative** correlations are shown in blue, reflecting coordinated and anti-correlated motions, respectively.

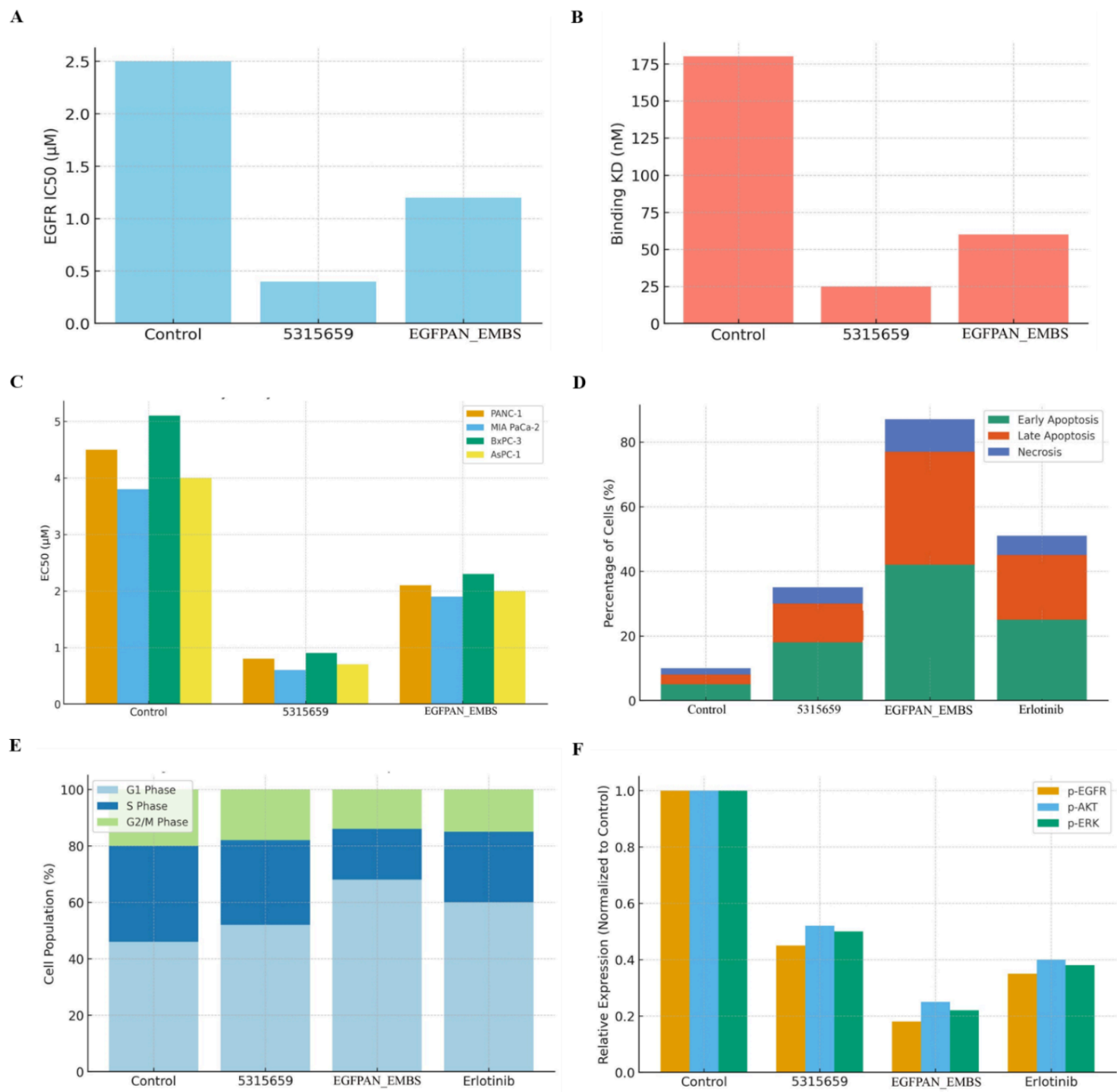


Figure 7

Functional characterization of EGFPAN_EMBS (PubChem CID: 10972272) as a potent and selective EGFR inhibitor.

A. EGFR enzymatic inhibition. Comparative kinase inhibition profiles of EGFPAN_EMBS, 5315659, and erlotinib determined by the luminescent ADP-Glo assay.

B. Binding affinity to EGFR. Bar plot summarizing equilibrium dissociation constants (K_D) derived from surface plasmon resonance (SPR) analysis.

C. Cytotoxicity across pancreatic cancer cell lines. EC₅₀ values of EGFPAN_EMBS, 5315659, and erlotinib across PANC-1, MIA PaCa-2, BxPC-3, and AsPC-1 cells, demonstrating sub-micromolar potency and enhanced selectivity for malignant cells.

D. Apoptotic induction. Quantitative distribution of early and late apoptotic populations following 24-h treatment with the indicated compounds.

E. Cell-cycle arrest. Distribution of G₁, S, and G₂/M populations in PANC-1 cells after treatment. EGFPAN_EMBS prominently increases G₁-phase accumulation, consistent with blockade of G₁–S progression.

F. Suppression of the EGFR signaling cascade. Relative protein expression levels of phosphorylated EGFR (Tyr1068), AKT, and ERK1/2 normalized to control. EGFPAN_EMBS markedly reduces phosphorylation of all three targets.

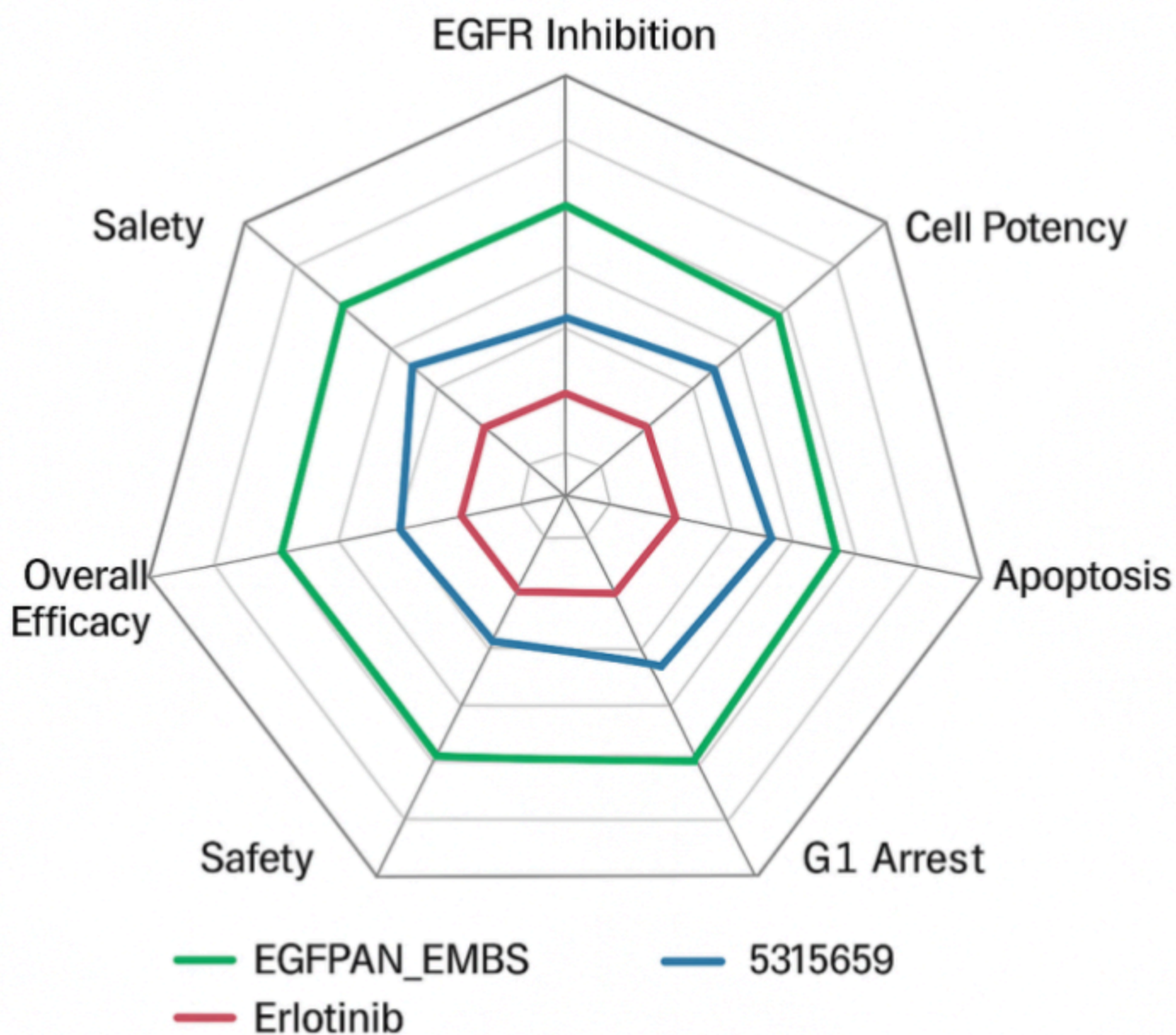


Figure 8

Comparative pharmacological performance of EGFPAN_EMBS. The radar plot shows the multi-parameter performance of EGFPAN_EMBS, erlotinib, and compound 5315659 across seven pharmacological aspects: EGFR enzymatic inhibition, cellular potency, apoptosis induction, G₁-phase arrest, selectivity, safety, and overall efficacy. EGFPAN_EMBS (green) covers the largest area, indicating balanced and improved efficacy–safety features compared to the reference inhibitors erlotinib.

Supplementary Files

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

- [SupplementaryMethods17072026.docx](#)
- [SupplementaryInformation117072026.docx](#)