

EGFR as a Potential Molecular Target of Cleistanthin A and Diphyllin Derivatives: An Integrated Computational and Biophysical Study

Bunnatut Panasawatwong

Mahidol University

Duangporn Lohawittayanan

Mahidol University

Napason Chabang

Mahidol University

Sitthivut Charoensutthivarakul

Mahidol University

Sakonwan Kuhaudomlarp

Mahidol University

Bamroong Munyoo

Mahidol University

Patoomratana Tuchinda

Mahidol University

Noppawan Phumala Morales

Mahidol University

Supachoke Mangmool

Chiang Mai University

Somrudee Reabroi

Mahidol University

Pimtip Sanvarinda

pimtip.san@mahidol.ac.th

Mahidol University

Bodee Nutho

Mahidol University

Article

Keywords: cleistanthin A, diphyllin, Epidermal Growth Factor Receptor (EGFR), molecular docking, computer-aided drug design (CADD), surface plasmon resonance (SPR)

Posted Date: August 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10332574/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 October 25th, 2026.
See the published version at <https://doi.org/10.1038/s41598-026-76701-8>.

EGFR as a Potential Molecular Target of Cleistanthin A and Diphyllin Derivatives: An Integrated Computational and Biophysical Study

Bunnatut Panasawatwong ¹, Duangporn Lohawittayanan ¹, Napason Chabang ²,
Sitthivut Charoensutthivarakul ^{1,3}, Sakonwan Kuhaudomlarp ⁴, Bamroong
Munyoo ³, Patoomratana Tuchinda ³, Noppawan Phumala Morales ¹, Supachoke
Mangmool ⁵, Somrudee Reabroi ¹, Pimtip Sanvarinda ^{1,*}, and Bodee Nutho ^{1,*}

¹ Department of Pharmacology, Faculty of Science, Mahidol University,
Bangkok, Thailand 10400

² School of Bioinnovation and Bio-based Product Intelligence, Faculty of
Science, Mahidol University, Bangkok, Thailand 10400

³ Excellence Center for Drug Discovery (ECDD), Faculty of Science,
Mahidol University, Bangkok, Thailand 10400

⁴ Department of Biochemistry, Faculty of Science, Mahidol University,
Bangkok, Thailand 10400

⁵ Department of Pharmaceutical Care, Faculty of Pharmacy, Chiang Mai
University, Chiang Mai, Thailand 50200

* Correspondence: pimtip.san@mahidol.ac.th and
bodee.nut@mahidol.ac.th

Abstract

Cleistanthin A (CTA) is an aryl-naphthalide lignan glycoside that exhibits anticancer activities through diverse molecular mechanisms. This study used an integrated computer-aided drug design (CADD) workflow, including target prediction, molecular docking, molecular dynamics (MD) simulations, binding free-energy calculations, and surface plasmon resonance (SPR), to identify a potential molecular target for CTA and its derivatives. SwissTargetPrediction identified protein kinases as potential molecular targets. Molecular docking against targetable non-small cell lung cancer-associated kinases revealed that CTA and its derivatives exhibited stronger binding to wild-type EGFR than erlotinib. MD simulations indicated the stability of the protein–ligand complexes, whereas Molecular Mechanics Poisson-Boltzmann (generalized Born) Surface Area (MM/PB(GB)SA) analyses identified van der Waals interactions as the primary driving force for binding at the ATP-binding pocket. Incorporation of a benzoyl ester moiety to this pharmacophore increased nonpolar binding energy contribution. SPR confirmed concentration-dependent binding of CTA, MUC-853, MUC-858, Compound 1, and Compound 2 to the EGFR kinase domain, with Compound 1 and MUC-858 exhibiting good dissociation constants (K_D) of 6.4 and 9.5 μM , respectively. Collectively, these findings denote aryl-naphthalide lignans as a potential scaffold for EGFR-targeted drug discovery and provide structure–interaction insights for rational design of inhibitors with improved potency and selectivity.

Keywords: cleistanthin A; diphyllin; Epidermal Growth Factor Receptor (EGFR); molecular docking; computer-aided drug design (CADD); surface plasmon resonance (SPR);

Introduction

The traditional drug discovery and development pipeline is labor-intensive, time-consuming, and costly, with approximately 90 percent of drug candidates failing during clinical development (1), particularly during late-stage clinical trials (2). In 2025, the U.S. Food and Drug Administration (FDA) announced plans to phase out animal testing requirements for many classes of drugs (3). Thus, computer-aided drug design (CADD) has emerged as an indispensable paradigm in modern medicinal chemistry and drug discovery. CADD provides rapid, cost-effective drug discovery, while offering personalized therapeutic insights (4). CADD enables high-throughput target identification, prediction of ligand binding modes, and assessment of the thermodynamic stability of protein–ligand complexes. Furthermore, it supports rapid virtual screening, structure-based hit identification, rational lead optimization, and mechanistic insights prior to experimental validation. By enriching high-quality candidates early in the discovery process, it reduces the need for extensive compound synthesis and minimizes reliance on extensive preliminary animal testing (5, 6).

Protein kinase dysregulation occurs in many cancers and contributes to aberrant cell proliferation, survival, motility, metabolism, and angiogenesis, all of which are recognized hallmarks of cancer (7). Consequently, protein kinases have become major therapeutic targets, and the development of kinase inhibitors has substantially transformed the treatment of numerous cancers (8, 9). Non-small cell lung cancer (NSCLC) is among the cancers that have benefited from kinase-targeted therapies (10). In NSCLC, the most frequently altered kinases include EGFR, ALK, BRAF, c-MET, ROS1, and RET (11). Targeting these oncogenic kinases has produced significant clinical benefits because tumors are highly dependent on their signaling activity for growth and survival (12). Among these, EGFR mutations are particularly prevalent in Asian patients with NSCLC, occurring in approximately 50% of cases (9). The common activating mutations, including exon 19 deletions and the L858R point mutation (7, 8) confer sensitivity to EGFR tyrosine kinase inhibitors (TKIs), such as erlotinib, gefitinib, and osimertinib.

Natural compounds have long been used in traditional medicine and continue to attract interest in drug discovery due to their structural diversity (13). Several plant-derived compounds, such as curcumin, quercetin, genistein, apigenin, capsaicin, and andrographolide, have been reported to exhibit kinase inhibitory activity (14, 15). However, natural compounds often act across multiple signaling pathways rather than at a single target, and most share

practical limitations including low potency, poor selectivity, and unfavorable pharmacokinetics (16). Chemical modification is therefore necessary to improve their drug-like properties (17). One effective strategy is halogenation: approximately 20% of small-molecule anticancer lead compounds contain fluorine, bromine, or chlorine. Halogens improve lipophilicity and membrane permeability, and their electronegativity can enhance binding affinity, metabolic stability, and target selectivity (18). Several studies have successfully applied CADD to identify kinase inhibitors derived from plant-based scaffolds (19-21).

Previous studies have shown that cleistanthin A (CTA), an aryl-naphthalide lignan glycoside isolated from *Cleistanthus collinus* or *Phyllanthus taxodiifolius* Beille, is cytotoxic against several human cancer cell lines (22, 23), including hepatocellular carcinoma, prostate, colorectal, breast, and lung cancer cell lines (23-26). Reported anticancer mechanisms include inhibition of DNA synthesis and induction of apoptosis (24, 27), inhibition of vacuolar-type H⁺-ATPase (V-ATPase), suppression of focal adhesion kinase (FAK) phosphorylation, and downregulation of the Wnt/ β -catenin pathway (24, 25, 28, 29), suggesting that CTA modulates multiple cancer-associated signaling pathways.

In the present study, we established an integrated multi-stage CADD workflow, followed by experimental validation using surface plasmon resonance (SPR), to identify the potential molecular targets of cleistanthin A (CTA) and diphyllin derivatives. The workflow incorporated target prediction, molecular docking, molecular dynamics (MD) simulations, and binding free-energy analyses to evaluate CTA and its derivatives, including benzoate ester derivatives and newly synthesized diphyllin analogs. The findings of this study help identify potential molecular targets for the aryl-naphthalide lignan scaffold and provide the atomic level of protein–ligand interaction insights that lay the foundation for the rational design and optimization of next-generation derivatives with enhanced potency and selectivity toward this target.

Results

Compounds synthesis

Cleistanthin A was acquired as previously described (23). MUC-853 and MUC-858 were synthesized in accordance with the previously reported protocol for related derivatives in this class (**Supplementary Method**) (30-32). A series of *O*-substituted diphyllin derivatives (1–5) was synthesized utilizing commercially obtained diphyllin as the core scaffolding. The target analogues were prepared through either etherification or esterification of the C-4 hydroxyl group (33). Compounds 1, 4, and 5 were synthesized via nucleophilic substitution reactions by treating diphyllin with the corresponding benzyl bromides in the presence of potassium carbonate (K_2CO_3) in dimethyl sulfoxide (DMSO) at 80 °C. Under similar alkylation conditions, the pyridinylmethyl analogue 3 was accessed using 3-(chloromethyl)pyridine in the presence of cesium carbonate (Cs_2CO_3) and a catalytic amount of potassium iodide (KI) in dimethylformamide (DMF) at 70 °C under a nitrogen atmosphere. Conversely, acylation of the C-4 position to generate the ester derivative 2 was accomplished using 3-(trifluoromethyl)benzoyl chloride in dichloromethane at room temperature, mediated by *N,N'*-diisopropylethylamine (DIPEA) and a catalytic quantity of 4-dimethylaminopyridine (DMAP). All target compounds were purified, affording the desired products. The chemical structures were confirmed by high-resolution ^1H - and ^{13}C -NMR spectroscopy.

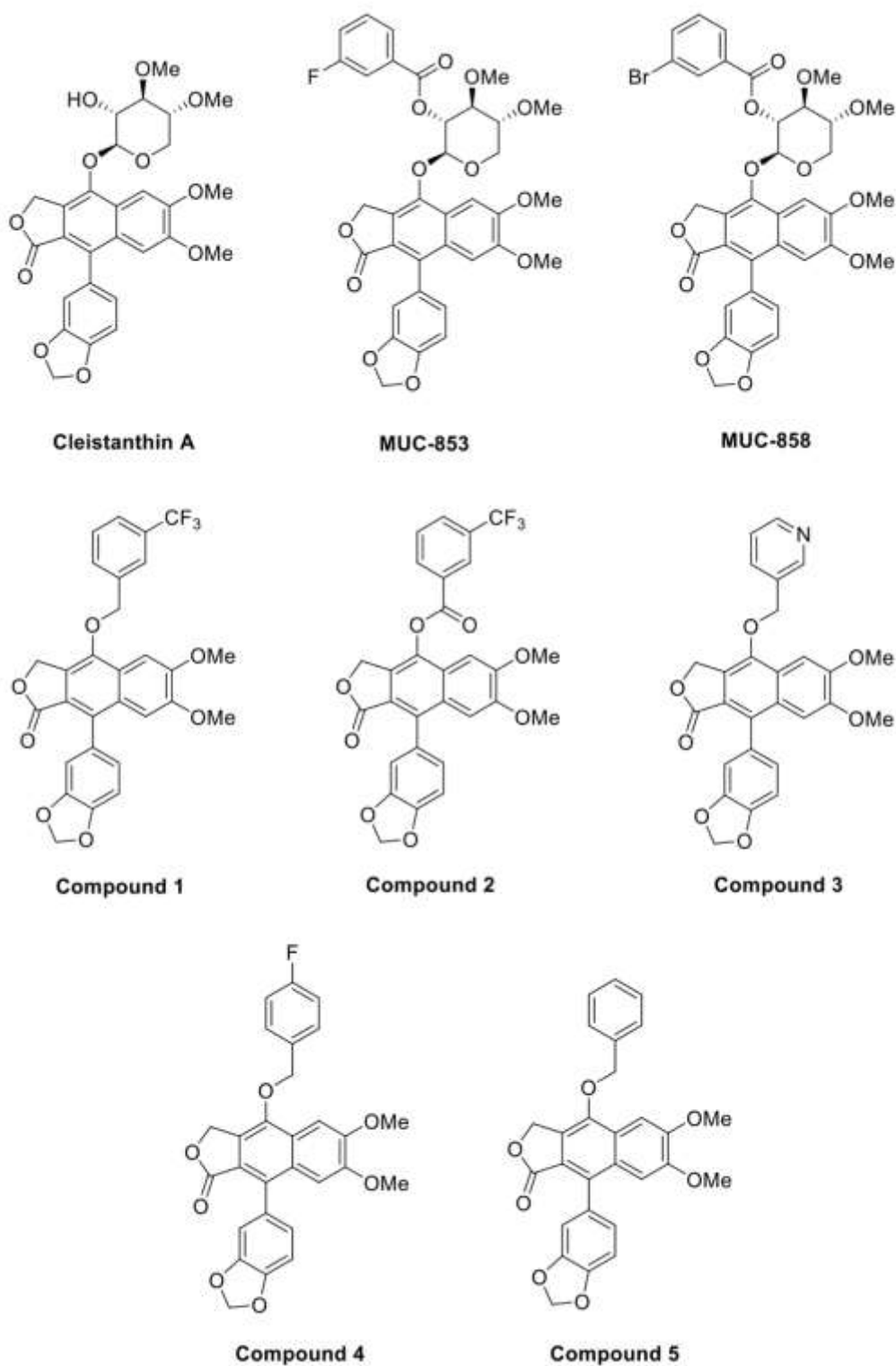


Figure 1. Chemical structures of cleistanthin A (CTA), MUC-853, MUC-858 and compounds 1-5

Molecular docking against key oncogenic protein kinases

Kinases are enzymes that catalyze the transfer of phosphate groups from ATP to substrate proteins, thereby regulating diverse cellular processes. Kinase domains are largely conserved across kinase families (34). Aberrant kinase signaling — arising from gene amplification or gain-of-function mutations in the kinase domain that confer constitutive activation — is strongly implicated in the pathogenesis of numerous cancers (35) and represent primary therapeutic targets in oncology (36). To test the probability of kinase interaction, the SwissTargetPrediction server (37) (accessed on 19 July 2024) was used to predict the potential molecular targets of CTA. Protein kinases were identified as one of the predicted target classes (**Table S4, supplementary data 1**). Based on the predicted binding possibility, kinases commonly mutated in lung cancer, including EGFR, ALK, BRAF, c-MET, ROS1, and RET (38) were selected for further screening.

Molecular docking was performed to predict the binding modes and molecular interactions of CTA and its derivatives within the kinase domain of each protein target, with approved kinase inhibitors serving as reference compounds. Binding energies were computed using standard molecular docking methodology. More negative values indicate stronger predicted binding affinity.

The calculated binding energies for CTA and its derivatives on the kinase proteins were presented in Figure S1. Only the targets for which the predicted binding energies of the test compounds surpassed those of the corresponding reference drugs were chosen for detailed analysis. Binding energy analysis against the TK domain of wild-type EGFR (WT-EGFR) suggested that all tested compounds exhibited superior predicted binding affinities relative to erlotinib (**Figure 2**). Therefore, EGFR was prioritized for further investigation. Docking simulations within the TK domain of WT-EGFR revealed that the tested compounds exhibited binding energies ranging from -8.213 to -11.22 kcal/mol, all of which outperformed erlotinib (-7.161 kcal/mol). Among the evaluated molecules, Compound 2 displayed the highest docking score (-11.22 kcal/mol), followed by Compound 1 (-11.09 kcal/mol), Compound 4 (-10.38 kcal/mol), Compound 5 (-10.295 kcal/mol), and Compound 3 (-10.074 kcal/mol). However, because standard molecular docking relies on a rigid protein structure, these docking scores may not fully reflect the binding behavior within a dynamic biological system. Therefore, molecular dynamics (MD) simulations were subsequently performed for the top-ranked compounds

(Compound 1-5, MUC-853, MUC-858), along with CTA and erlotinib, to evaluate the complex stability over time.

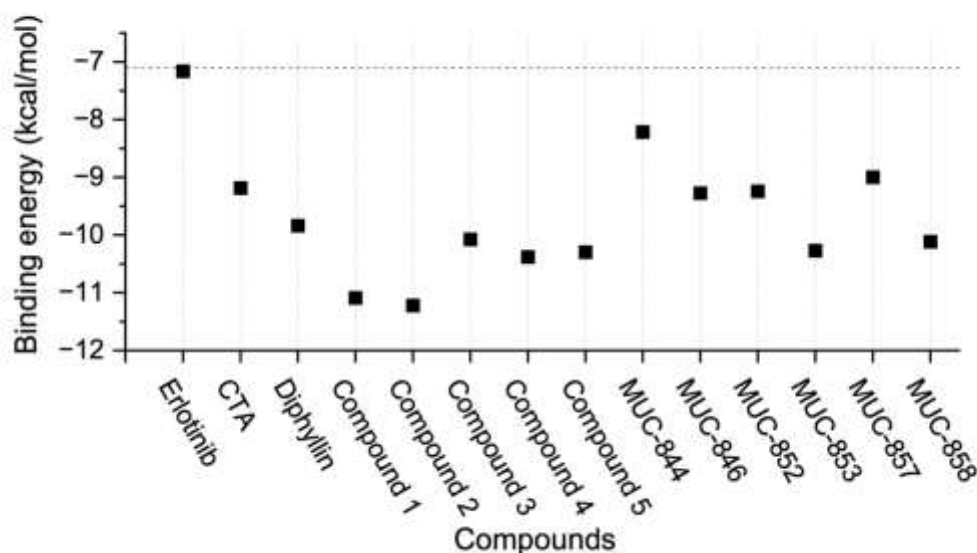
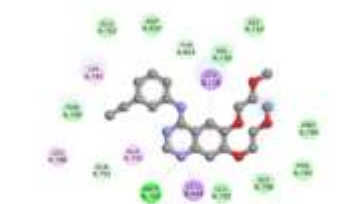


Figure 2. Docking scores of CTA and derivatives against kinase domain of WT-EGFR compared to the reference drug, erlotinib.

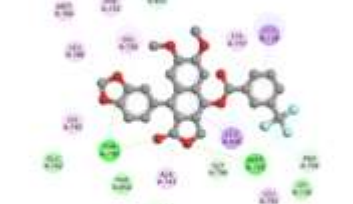
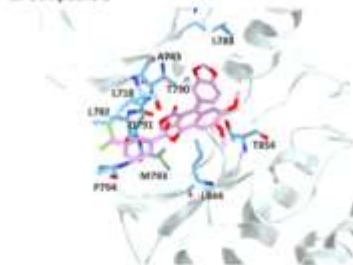
Further structural analysis provided several key insights into the structure-interaction relationship of these derivatives. First, substitution of the free hydroxyl group on the diphyllin scaffold with a linker bearing a cyclic moiety conferred improved binding affinity relative to the parent compound, diphyllin (−9.838 kcal/mol). This enhancement is likely attributable to supplementary hydrophobic interactions, specifically pi-sigma or pi-alkyl bonds between the cyclic moiety and Leu718 (**Figure 3**). Introduction of the halobenzoate group into the glycosidic moiety of CTA likewise enhanced binding affinity, as observed for MUC-853 (−10.271 kcal/mol) and MUC-858 (−10.115 kcal/mol) relative to the parent compound CTA (−9.182 kcal/mol). These gains are similarly attributed to the formation of additional halogen bonds. The aryl naphthalene core structure conserved among all derivatives is responsible for interaction with the TK domain residues, including Leu718, Val726, Ala743, Met793, and Leu844 as essential residues with interaction detected in most EGFR kinase inhibitors (39). Furthermore, the test compounds can interact with Lys745, a critical residue for salt bridge formation during kinase catalysis (40), and Thr790, the well-characterized gatekeeper residue

whose mutation confers resistance to first- and second-generation inhibitors (41). The interaction profile of this scaffold was closely related to the binding profile of classical EGFR kinase inhibitors, including quinazoline-based compounds (42) and pyrimidines/pyridines-based derivatives (43).

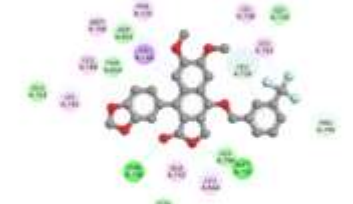
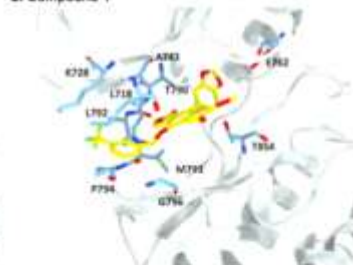
a. Erlotinib



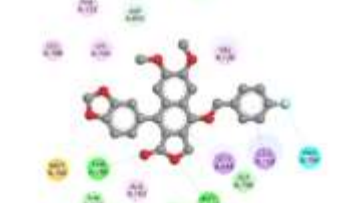
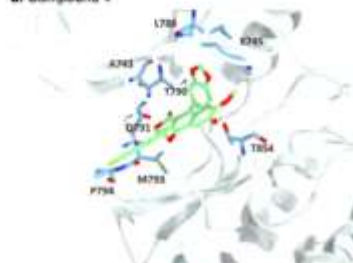
b. Compound 2



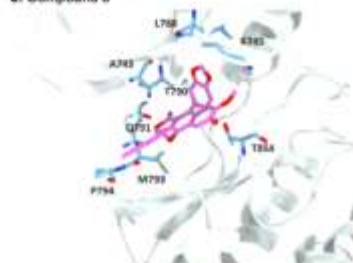
c. Compound 1



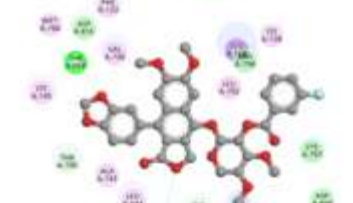
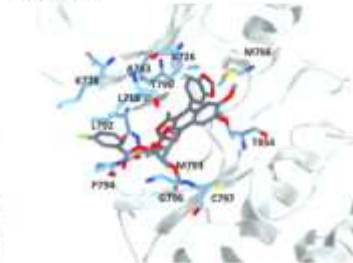
d. Compound 4



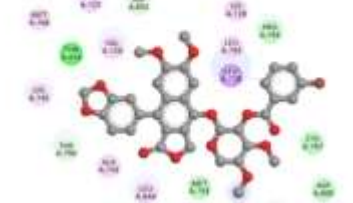
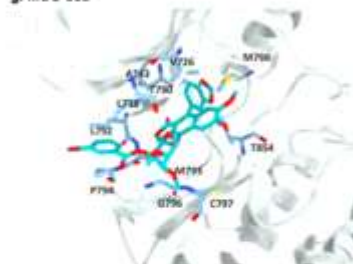
e. Compound 5



f. MUC-853



g. MUC-858



h. Compound 3

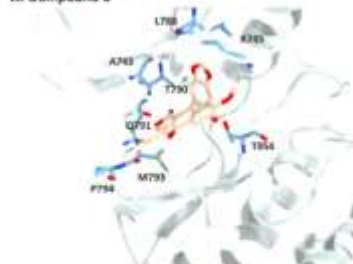


Figure 3. 2D interaction profiles within the ATP-binding pocket of WT EGFR. **a.** erlotinib (control) and the top-ranked compounds in order of highest to lowest docking score: **b.** Compound 2, **c.** Compound 1, **d.** Compound 4, **e.** Compound 5, and **f.** MUC-853, **g.** MUC-858, **h.** Compound 3, and **i.** CTA.

Complexes stability assessed by molecular dynamics (MD) simulation

Molecular dynamics (MD) simulations were performed in an aqueous environment to evaluate the structural characteristics and stability of the protein–ligand complexes. Three key structural parameters were monitored throughout the trajectories to assess system stability: root-mean-square deviation (RMSD), radius of gyration (R_g), and the number of intermolecular hydrogen bonds (**Figure 4**). The RMSD profiles of the complex atoms exhibited minor fluctuations of 2–4 Å during the initial 100 ns of simulation, with average RMSD values ranging from 2.7 to 3.7 Å and standard deviations between 0.4 and 1.0 Å. These initial variations reflected the expected structural adjustments of the protein upon ligand binding. Subsequently, these fluctuations subsided and exhibited mean RMSD values between 2.6 and 4.3 Å with lower standard deviations (0.1–0.6 Å), indicating that the majority of the protein–ligand complexes had converged toward relatively stable state. Compound 4, Compound 5, and MUC-858 retained slightly higher fluctuations throughout the trajectory; this behavior indicated a higher degree of localized conformational flexibility within these complexes rather than any large-scale structural transitions or destabilization.

To determine the overall structural compactness of the protein upon ligand binding, R_g values were computed using the protein's alpha-carbon (C_α) atoms. Across all simulated systems, the R_g values remained steady, ranging between 19.20 Å and 20.99 Å. This narrow distribution indicated that the protein maintained a highly rigid and compact conformation without undergoing any discernible unfolding. Finally, an analysis of intermolecular hydrogen bonding profiles indicated that CTA and its derivatives displayed a high number of interactions, consistently maintaining 2 - 7 hydrogen bonds. This enhanced hydrogen-bonding network likely contributes favorably to their overall protein–ligand binding affinity. Collectively, these structural analyses indicated that all simulation models achieved a high degree of relative stability.

Consequently, the structural coordinates from the final 50 ns of the trajectories were adopted as the production trajectory for subsequent analyses.

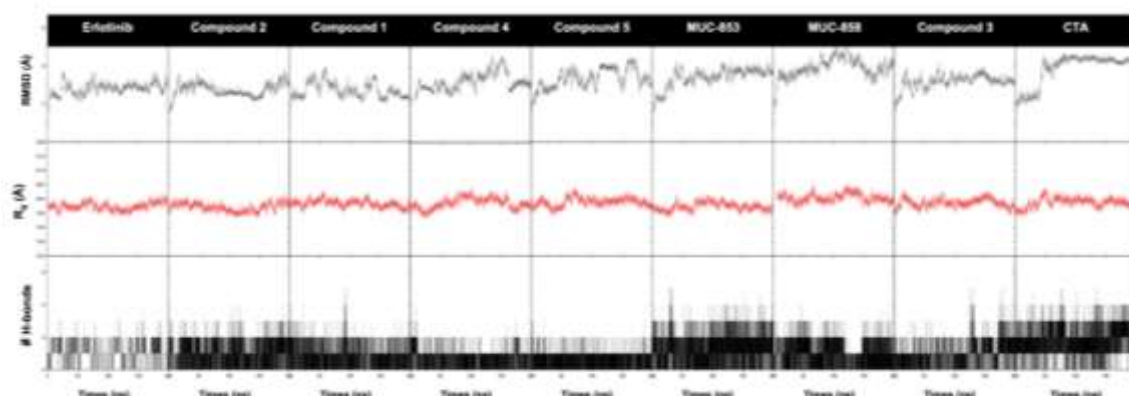


Figure 4. Time evolutions of MD structural stability parameters over 200 ns of CTA and its derivatives in complex with kinase domain of WT-EGFR. The graphs display the time evolution of the root-mean-square deviation (RMSD), the radius of gyration (R_g) of $C\alpha$ atoms, and the number of intermolecular hydrogen bonds (# H-bonds) during 200 ns.

Binding affinity of protein-ligand complexes

Binding free energies were estimated from the final 50 ns of equilibrated trajectory to elucidate the thermodynamic forces governing ligand association with the WT-EGFR kinase domain, employing the Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics Generalized Born Surface Area (MM/GBSA) formalisms.

The decomposed energy components and total binding free energies (ΔG_{total}) of the test compounds and erlotinib are summarized in **Table 1**. Both methods yielded consistent compound rankings across systems. MM/GBSA calculations produced slightly more favorable (more negative) binding free energies than MM/PBSA, attributable to the less stringent polar solvation penalties intrinsic to the GB approximation (44).

Among all evaluated compounds, the halobenzoate glycoside derivatives exhibited the most favorable thermodynamic binding profiles. Specifically, MUC-858 exhibited the most favorable binding free energy ($\Delta G_{\text{total}} = -47.06$ kcal/mol via MM/GBSA; -31.91 kcal/mol via MM/PBSA), closely followed by

MUC-853 ($\Delta G_{\text{total}} = -46.71$ kcal/mol via MM/GBSA; -30.37 kcal/mol via MM/PBSA). Both derivatives exhibited superior binding affinities relative to the unmodified parent compound CTA and erlotinib. Decomposition of the gas-phase energy interactions (ΔE_{MM}) revealed that van der Waals interactions (ΔE_{vdW}) served as the primary driving force for complex stabilization across all systems. The magnitude of ΔE_{vdW} was approximately 2–4 times greater than that of the polar electrostatic interactions (ΔE_{ele}). MUC-853 and MUC-858 also exhibited the highest van der Waals contributions, at -66.67 kcal/mol and -66.19 kcal/mol, respectively. Therefore, van der Waals has been identified as the primary driving force for complex formation within the ATP-binding pocket of EGFR, consistent with reports for several EGFR inhibitors such as erlotinib, afatinib, Osimertinib (45), quinoxalinones, and fuopyridine derivatives (43, 46). In contrast, the total solvation free energy (ΔG_{sol}) offset the favorable gas-phase interactions, thereby reducing the overall binding free energy. The polar solvation term ($\Delta G_{\text{sol}}^{\text{ele}}$) contributed an unfavorable positive value that was 5–7 times stronger than the favorable non-polar solvation effect ($\Delta G_{\text{sol}}^{\text{nonpolar}}$), thereby acting as the main thermodynamic penalty to binding.

The diphyllin derivatives featuring a linker attached to a cyclic moiety (Compound 1 - 5) exhibited weaker binding affinities than erlotinib, displaying MM/GBSA values ranging from -28.84 to -35.93 kcal/mol and MM/PBSA values from -21.30 to -22.95 kcal/mol. Their binding was consistently mediated by non-polar interactions ($\Delta E_{\text{vdW}} + \Delta G_{\text{sol}}^{\text{nonpolar}}$) rather than polar terms ($\Delta E_{\text{ele}} + \Delta G_{\text{sol}}^{\text{ele}}$). Additionally, the halobenzoate glycoside derivatives displayed higher favorable total nonpolar binding energy compared to CTA. Given the predominantly lipophilic character of the EGFR ATP-binding pocket, the rational incorporation of lipophilic halogenated moieties represents a highly promising optimization pathway for scaffold optimization.

Table 1. Total binding free energy and thermodynamic energy components (kcal/mol) of CTA and derivatives calculated via MM/PBSA and MM/GBSA methods. Data are presented as mean $\pm$ SEM

Energy component	Erlotinib	Compound 2	Compound 1	Compound 4	Compound 5	MUC-853	MUC-858	Compound 3	CTA
Gas term									
ΔE_{vdW}	-46.6 $\pm$ 0.13	-42.73 $\pm$ 0.16	-51.85 $\pm$ 0.13	-50.23 $\pm$ 0.13	-49.48 $\pm$ 0.12	-66.67 $\pm$ 0.14	-66.19 $\pm$ 0.15	-48.47 $\pm$ 0.27	-50.93 $\pm$ 0.15
ΔE_{ele}	-17.27 $\pm$ 0.27	-22.25 $\pm$ 0.18	-20.01 $\pm$ 0.26	-13.32 $\pm$ 0.21	-13.65 $\pm$ 0.17	-19.17 $\pm$ 0.22	-15.58 $\pm$ 0.22	-13.94 $\pm$ 0.26	-23.24 $\pm$ 0.27
ΔE_{MM}	-63.87 $\pm$ 0.33	-64.98 $\pm$ 0.25	-71.86 $\pm$ 0.28	-63.56 $\pm$ 0.22	-63.12 $\pm$ 0.2	-85.84 $\pm$ 0.27	-81.77 $\pm$ 0.27	-62.41 $\pm$ 0.32	-74.17 $\pm$ 0.31
Solvation term									
PBSA									
$\Delta G_{sol(PBSA)}^{ele}$	41.87 $\pm$ 0.33	48.17 $\pm$ 0.26	54.37 $\pm$ 0.34	45.96 $\pm$ 0.28	46.64 $\pm$ 0.25	61.29 $\pm$ 0.3	55.95 $\pm$ 0.33	44.12 $\pm$ 0.33	50.68 $\pm$ 0.25
$\Delta G_{sol(PBSA)}^{nonpolar}$	-4.63 $\pm$ 0.01	-4.55 $\pm$ 0.01	-5.07 $\pm$ 0.01	-4.82 $\pm$ 0.01	-4.81 $\pm$ 0.01	-5.83 $\pm$ 0.01	-6.09 $\pm$ 0.01	-4.67 $\pm$ 0.02	-4.85 $\pm$ 0.01
$G_{sol(PBSA)}$	37.24 $\pm$ 0.33	43.63 $\pm$ 0.25	49.29 $\pm$ 0.34	41.14 $\pm$ 0.28	41.82 $\pm$ 0.25	55.47 $\pm$ 0.3	49.86 $\pm$ 0.33	39.45 $\pm$ 0.32	45.83 $\pm$ 0.25
GBSA									
$\Delta G_{sol(GBSA)}^{ele}$	36.63 $\pm$ 0.26	41.81 $\pm$ 0.19	42.91 $\pm$ 0.26	34.66 $\pm$ 0.2	34.43 $\pm$ 0.15	47.3 $\pm$ 0.21	42.73 $\pm$ 0.21	34.73 $\pm$ 0.22	44.65 $\pm$ 0.23
$\Delta G_{sol(GBSA)}^{nonpolar}$	-6.17 $\pm$ 0.02	-5.67 $\pm$ 0.02	-6.97 $\pm$ 0.02	-6.48 $\pm$ 0.01	-6.41 $\pm$ 0.01	-8.16 $\pm$ 0.01	-8.02 $\pm$ 0.02	-6.02 $\pm$ 0.04	-6.61 $\pm$ 0.01
$\Delta G_{sol(GBSA)}$	30.45 $\pm$ 0.25	36.14 $\pm$ 0.18	35.93 $\pm$ 0.25	28.18 $\pm$ 0.2	28.02 $\pm$ 0.14	39.13 $\pm$ 0.21	34.71 $\pm$ 0.2	28.71 $\pm$ 0.22	38.04 $\pm$ 0.22
Binding free energy									
$\Delta G_{total(MM/PBSA)}$	-26.64 $\pm$ 0.16	-21.35 $\pm$ 0.16	-22.56 $\pm$ 0.21	-22.42 $\pm$ 0.2	-21.3 $\pm$ 0.21	-30.37 $\pm$ 0.22	-31.91 $\pm$ 0.24	-22.95 $\pm$ 0.21	-28.34 $\pm$ 0.2
$\Delta G_{total(MM/GBSA)}$	-33.42 $\pm$ 0.15	-28.84 $\pm$ 0.14	-35.93 $\pm$ 0.12	-35.37 $\pm$ 0.12	-35.1 $\pm$ 0.12	-46.71 $\pm$ 0.14	-47.06 $\pm$ 0.15	-33.7 $\pm$ 0.22	-36.13 $\pm$ 0.16

Per-residue Free Energy Decomposition and Hot-Spot Residue Identification

To understand the key amino acids in the ATP-binding pocket of WT-EGFR kinase domain that are responsible for ligand interaction, per-residue free energy decomposition was performed using the MM/GBSA method. The analysis represented the energy contributed to each amino acid against the compound complex, with a threshold of ≤ -1.0 kcal/mol to specify important binding hotspots (**Figure 5**).

Erlotinib displayed critical interactions with Leu718 (-2.34 kcal/mol) in the phosphate-binding loop (P-loop), Val726 (-1.68 kcal/mol) in the hydrophobic ATP-binding pocket, Leu792 (-1.67 kcal/mol) and Met793 (-1.97 kcal/mol) in the hinge region. This binding profile is in strong agreement with those reported for other established EGFR-TK inhibitors, including gefitinib, osimertinib, quinoxalinones, anilino-1,4-naphthoquinones, thieno[2,3-d]pyrimidines, and furopyridines (45-49).

CTA and derivatives exhibited a binding pattern broadly similar to that of erlotinib, with greater energy contribution at Leu718, Val726 and Leu792. Specifically, MUC-853 and MUC-858 showed highly favorable energy contributions within the hydrophobic pocket at Leu718 (-3.47 and -3.58 kcal/mol, respectively) and Val726 (-3.26 and -3.18 kcal/mol, respectively). These results are consistent with the enhanced non-polar energy contributions mentioned earlier. An additional interaction at Phe723 was observed exclusively in MUC-853 (-1.56 kcal/mol) and MUC-858 (-2.09 kcal/mol), suggesting they adopt a binding conformation that enables supplementary interactions (e.g., a π -alkyl interaction). However, the energy contribution at Met793, a residue whose engagement has been correlated with kinase inhibitory potency (50), was lower for CTA and its derivatives than for erlotinib. Compound 1, Compound 4, and Compound 5 contain flexible linker with a cyclic moiety and displayed additional interaction with Lys745, which is important for salt bridge formation during the kinase catalytic reaction. The percentage of hydrogen bond occupancy of all complexes was calculated and demonstrated in **Figure 5**. This analysis was based on two geometric criteria: (i) an acceptor-to-donor ($A\cdots D$) distance of ≤ 3.5 Å, and (ii) an acceptor-hydrogen-donor ($A\cdots H-D$) angle of $\geq 120^\circ$ (51).

MD trajectory analysis revealed that the reference inhibitor erlotinib formed two principal hydrogen bonds (H-bonds) within the ATP binding pocket. The dominant interaction was between the nitrogen atom of the

quinazoline core and the backbone amine hydrogen of Met793, in which the quinazoline nitrogen acted as a hydrogen bond acceptor and maintained a 98% occupancy throughout the simulation. A second H-bond was observed between an oxygen atom of erlotinib and Cys797, with a 25% occupancy.

Similarly, CTA and its derivatives formed H-bonds with the hinge residue Met793 through the oxygen atoms of the lactone ring, which functioned as hydrogen bond acceptors. The occupancy of these interactions ranged from 14% to 84%, indicating varying degrees of interaction stability among the compounds. The compounds also formed H-bonds with other key residues in the ATP-binding pocket. For instance, CTA interacted with Thr790, with a hydrogen bond occupancy of 31%, whereas Compound 2 formed an additional H-bond with Cys797 (12% occupancy). These supplementary interactions provide structural insights into the distinct binding modes of CTA derivatives and offer a rational basis for further optimization of EGFR-TK binding affinity.

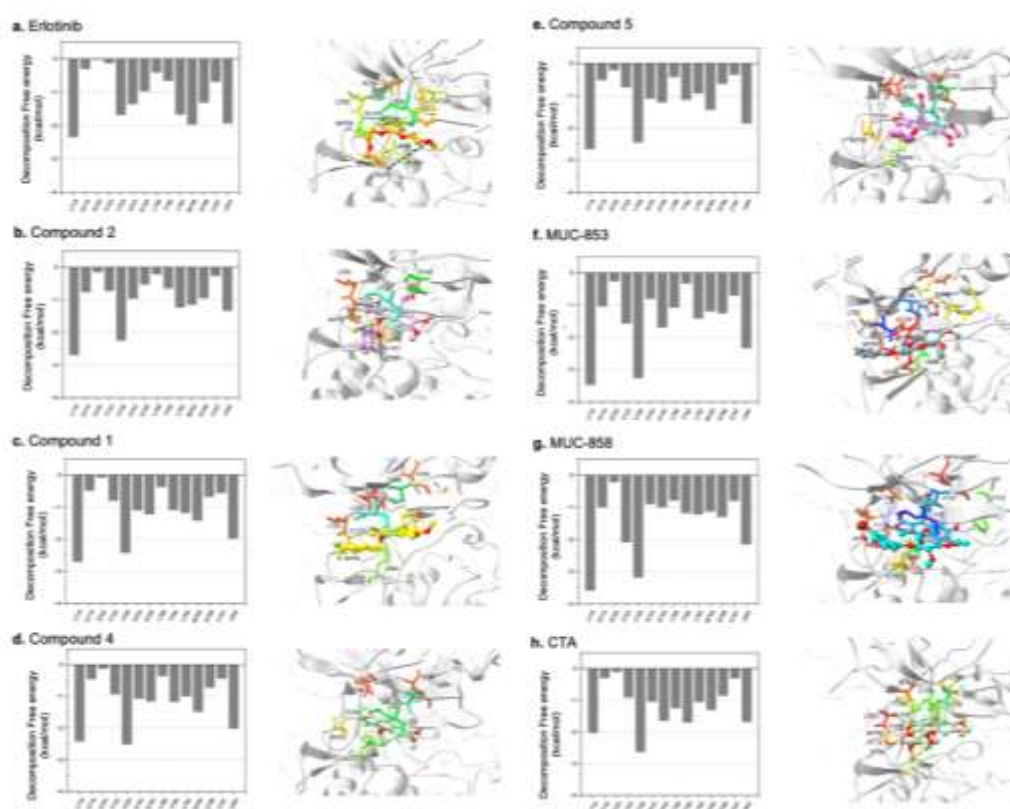


Figure 5. Per-residue free energy decomposition calculated with the MM/GBSA method. a. Erlotinib, b. Compound 2, c. Compound 1, d. Compound 4, e. Compound 5, f. MUC-853, g. MUC-858, and h. CTA. The bar plot (left) shows key residues contributing to ligand binding, with threshold of

≤ -1.0 kcal/mol. The 3D binding orientation (right) is captured from the final MD snapshot, with key residues colored according to energy contributions ranging from -3.5 (blue) to -1.0 kcal/mol (red). Specific acceptor-donor hydrogen bonds are shown as black dashed lines labeled with percentage of occurrence during the final 50 ns of the MD simulation.

Surface plasmon resonance

To confirm the physical binding of test compounds to EGFR protein, surface plasmon resonance (SPR) was employed. Compounds 4 and 5 were insoluble in dimethyl sulfoxide (DMSO), which was used as the standard solvent, at the desired concentration of 10 mM and were therefore excluded from the study. Eight compounds including CTA, MUC-853, MUC-858, compound 1, compound 2, compound 3, erlotinib, and osimertinib, respectively, were injected over the Ni-captured surface to test the non-specific binding to nickel ion on the NTA surface. The non-specific-binding control cycles confirmed no measurable analyte response over the nickel–NTA surface in the absence of captured EGFR (data not shown), confirming that all binding signals reported herein are attributable to the immobilized EGFR.

An initial screen at 1 and 10 μ M was conducted to assess dose-dependent binding responses. Of the eight compounds tested, six compounds including CTA, MUC-853, MUC-858, Compound 1, Compound 2, and the control inhibitor, osimertinib, produced a clear dose-dependent response above the buffer-blank baseline (**Figure 6**) and were advanced to the full affinity titration.

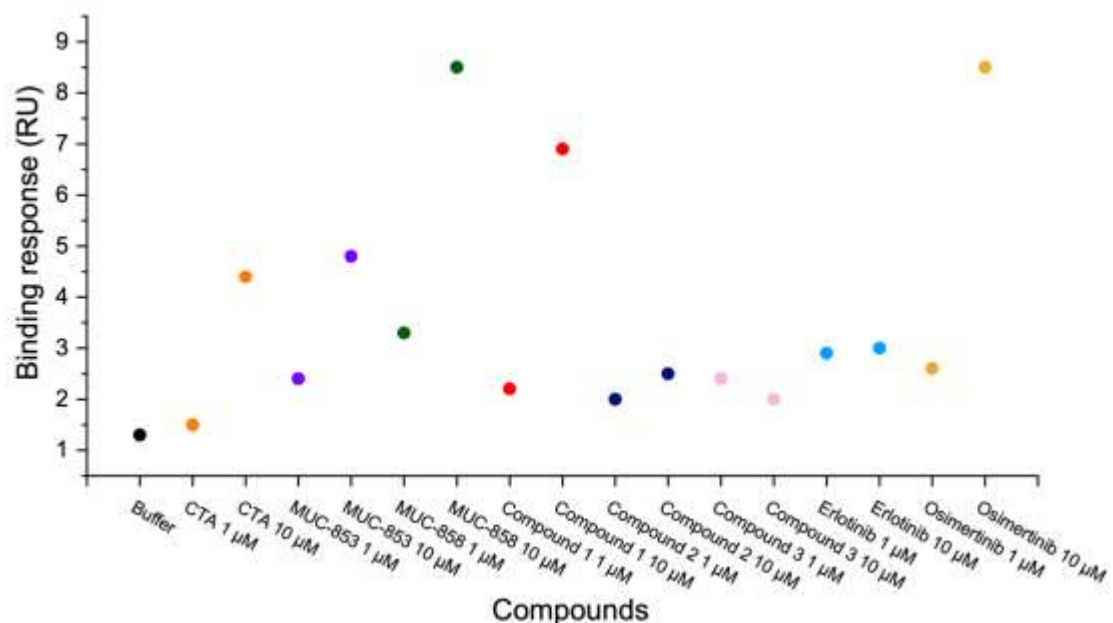


Figure 6. Binding response from initial screens. The compounds were injected in consecutive cycles at 1 and 10 μ M. The following compounds CTA, 853, 858, Compound 1, Compound 2, Compound 3 Erlotinib and ostimertinib were injected sequentially.

The five-point analyte series (25, 50, 75, 125 and 250 μ M) was injected over EGFR-loaded surfaces for each carried-forward compound. CTA, MUC-853, MUC-858, Compound 1, and Compound 2 produced concentration-dependent positive binding responses (**Figure 6**). Compound 1, MUC-853, and MUC-858 could be fitted with a steady-state single-site affinity model, giving K_D values in 6.4 – 34.6 μ M range (**Figure 7**). The remaining compounds, CTA and Compound 2, produced sensorgrams in which the response was negative or behaved anomalously across the concentration series and could not be fit to a binding model (**Figure S2 – Supplementary data 1**), therefore the K_D of these compounds not be determined. Compound 2 reflected non-specific interaction to NTA surface on flow cell 1 which overweight the binding response with the immobilized EGFR on flow cell 2.

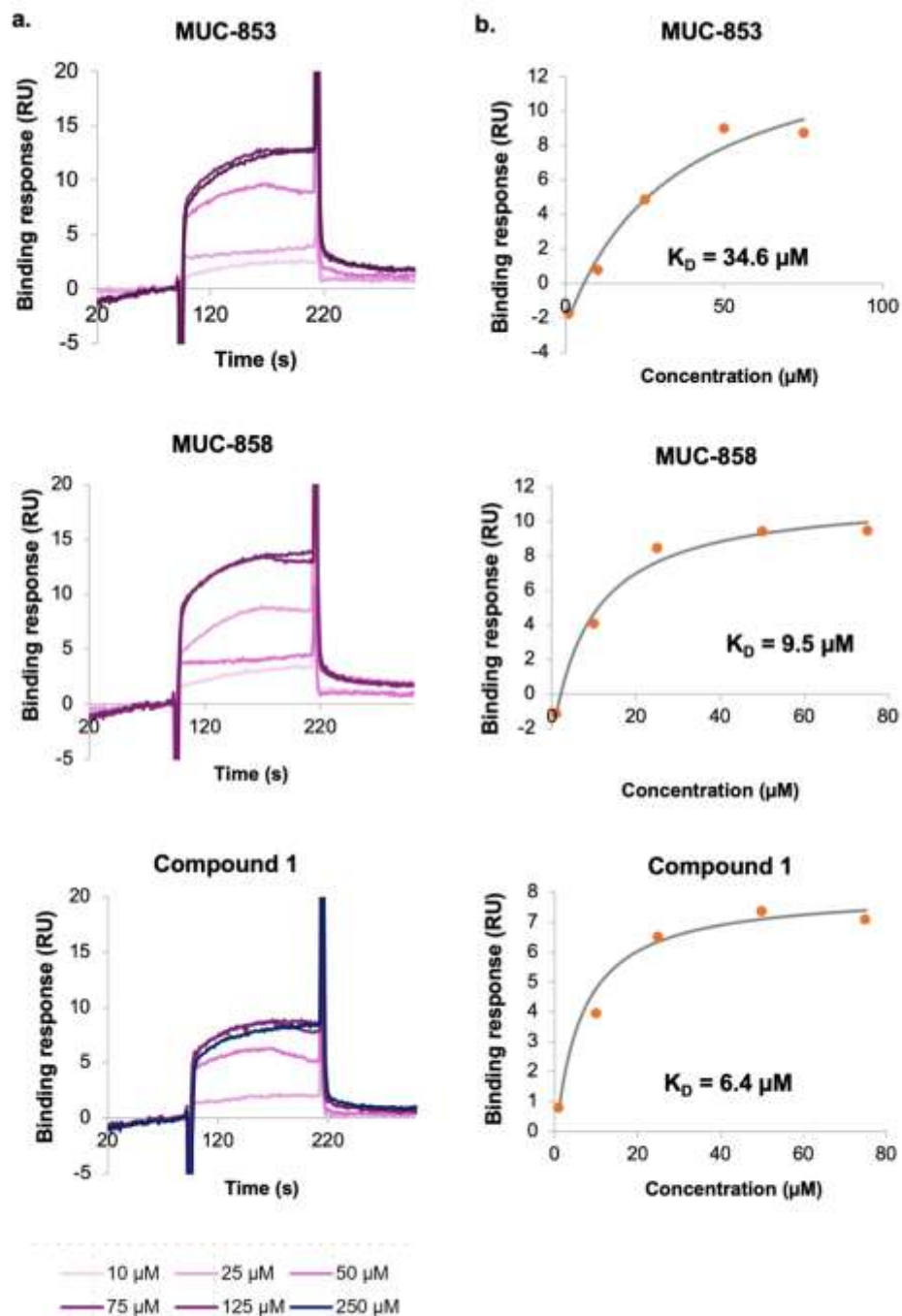


Figure 7. Steady state affinity analysis. a. Sensorgram shows the binding responses when injecting analyte with 5 different concentrations. **b.** Steady-state analyses.

Discussion

Cleistanthin A (CTA) is an aryl-naphthalide lignan glycoside isolated from the tropical plant *Cleistanthus collinus* (22) and *Phyllanthus taxodiifolius* (23). Structurally, CTA is a glycosylated derivative of diphyllin. Both CTA and its aglycone, diphyllin, have been reported to possess anticancer activity. Early studies demonstrated that CTA exhibits cytotoxicity against a variety of cancer cell lines, including murine leukemia (P-388), head and neck cancer (52), cervical cancer (SiHa), colon cancer (Col-2, HTC-116, SW480) (24), breast cancer (MCF-7), and lung cancer (Lu-1 and A549) (22, 23, 26). CTA exerts its anticancer activity through multiple mechanisms, including inhibition of DNA synthesis and induction of apoptosis (24, 27). Among its proposed molecular targets, vacuolar H⁺-ATPase (V-ATPase) has been identified as a key mediator of its biological activity (28, 29). In addition, the CTA derivative ECDD-18 has been reported to induce cytotoxicity by inhibiting V-ATPase, impairing autolysosome and interrupting autophagic flux (52), suggesting that the compound may modulate multiple cancer-associated signaling pathways.

In this study, we combined computer-aided drug design (CADD) with bioinformatic target-prediction databases to identify a molecular target for cleistanthin A and a series of newly synthesized diphyllin derivatives (**Figure 1**), and identified the epidermal growth factor receptor (EGFR) as a promising candidate. We first used the SwissTargetPrediction (37) server to predict potential targets of the compounds, which highlighted the protein kinase family. To validate the target prediction, molecular docking was performed against kinase proteins commonly mutated in lung cancer, and the predicted binding energies were compared with those of clinically approved inhibitors. We found that both wild-type EGFR (**Figure 2**) and the mutant forms (**Figure S1; G–I**, **supplementary data 1**) displayed binding energies superior to the reference inhibitors. Although docking and related methods are computationally efficient and well suited to predicting binding modes and distinguishing binders from non-binders, their quantitative accuracy is limited, and they generally cannot resolve differences in binding affinity smaller than one order of magnitude ($\Delta G_{\text{bind}} < 6 \text{ kJ/mol}$ or 1.43 kcal/mol) (53).

To further evaluate the stability of the ligand–protein complexes, we monitored the RMSD, R_g , and the number of intermolecular hydrogen bonds over the course of the simulations. The complex RMSD profiles showed minor fluctuations of 2–4 Å during the first 100 ns and subsequently stabilized in all systems. The R_g values remained steady, ranging from 19.20 to 20.99 Å, indicating that the complexes retained a compact conformation (**Figure 4**).

Additionally, the intermolecular hydrogen-bonding profiles showed that CTA and its derivatives formed a consistently high number of contacts, maintaining between 2 to 7 hydrogen bonds throughout the trajectories. We further applied the MM/PB(GB)SA methods (54, 55) to estimate the binding free energies between the compounds and EGFR. Decomposition of the gas-phase interaction energy revealed that van der Waals interactions were the primary driving force for complex stabilization across all systems. The diphyllin derivatives bearing a linker attached to a cyclic moiety (Compounds 1–5) showed weaker binding affinities than erlotinib. Despite their lower overall affinities, their binding remained predominantly mediated by non-polar interactions rather than polar terms. Within the halobenzoate glycoside series, nearly all structural modifications increased the total non-polar binding energy relative to CTA, indicating that incorporation of a halobenzoate group as a lipophilic substituent is an effective strategy for enhancing binding affinity, consistent with the predominantly lipophilic character of the EGFR ATP-binding pocket (56). Per-residue free-energy decomposition and hot-spot analysis showed that CTA and its derivatives shared a binding pattern broadly similar to that of erlotinib, with the strong energy contributions from the residues Leu718, Val726, and Leu792. Taken together, these results support a favorable interaction between the compounds and EGFR-TK.

Finally, to confirm direct physical binding, we employed surface plasmon resonance (SPR). In the initial screen, CTA, MUC-853, MUC-858, Compound 1, and Compound 2 produced dose-dependent responses and were selected for detailed analysis (**Figure 6**). Of these, MUC-853, MUC-858, Compound 1, and Compound 2 produced concentration-dependent, positive sensorgrams that could be fitted with a steady-state single-site affinity model. Among all compounds screened, Compound 1 and MUC-858 were the most promising, with K_D values of 6.4 and 9.5 μM , respectively (**Figure 7, Table 2**). Together, these findings establish a CADD-driven workflow for hit-to-lead screening against EGFR. Further work is needed to characterize the kinase-inhibitory activity of these compounds and their binding to the more clinically relevant mutant receptors, including EGFR L858R, EGFR L858R/T790M, and EGFR L858R/T790M/C797S (57) (**Supplementary data 2**). Moreover, using the structural insights into core-scaffold binding to guide the synthesis of new derivatives should support the development of novel aryl naphthalide-based kinase inhibitors. Nonetheless, the present study is limited to computational and biophysical analyses. Therefore, further validation using cell-based and in vivo models is required to confirm the biological relevance of these findings.

Conclusion

In this study, we established a CADD-driven workflow to identify potential targets of cleistanthin A (CTA) and novel diphyllin derivatives. Molecular docking, MD simulations, and binding free energy calculations revealed that CTA and its derivatives engage key EGFR residues (Leu718, Val726, Leu792) through predominantly van der Waals interactions, mirroring the binding mode of erlotinib. Incorporating a halobenzoate group enhances hydrophobic contacts within the EGFR ATP-binding pocket, and improves overall binding affinity. Surface plasmon resonance confirmed direct binding to EGFR, with Compound 1 and MUC-858 giving the lowest dissociation constants among the compounds tested. These results support the aryl-naphthalide lignan scaffold as a promising foundation for the rational design of next-generation EGFR inhibitors with improved potency and selectivity against clinically relevant EGFR mutants.

Materials and methods

Chemicals and reagents

All reagents and solvents were purchased from commercial suppliers (Sigma-Aldrich, Merck, or Tokyo Chemical Industry (TCI)) and were used without further purification. Diphyllin was purchased from Chembridge (Cat No. #5468327) (San Diego, USA). Reaction progress and preliminary purity assessment were monitored by thin-layer chromatography (TLC) using pre-coated silica gel 60 PF254 aluminum sheets (20 x 20 cm, layer thickness 0.2 mm) and the spots were visualized under UV light at 254 and 366 nm. High-resolution nuclear magnetic resonance (NMR) spectra were recorded on either a Bruker Ascend™ 300, Bruker Ascend™ 400 or a Bruker AVANCE-500 spectrometer at the Department of Chemistry, Faculty of Science, Mahidol University in CDCl₃ using tetramethylsilane (TMS) as an internal standard. Chemical shifts (δ) are reported in parts per million (ppm) and coupling constants (J) are given in Hertz (Hz). Purification of the compounds was carried out using appropriate chromatographic techniques, including silica gel column chromatography, preparative thin-layer chromatography, and automated column chromatography using a Biotage® Selekt system.

His-tagged recombinant EGFR was obtained from MedChemExpress (Cat. No. HY-P72987, lot 1101706). A Sensor Chip NTA was purchased from Cytiva, USA.

Isolation of Cleistanthin A and synthesis of its halobenzoate derivatives; MUC 853 and MUC 858

Cleistanthin A was extracted and isolated according to the procedure described by Tuchinda *et al.* (2006) (23). *Phyllanthus taxodiifolius* was collected in Amnartchareon Province, Thailand, in 2002 and deposited at the Forest Herbarium, Ministry of Natural Resources and Environment, Bangkok (voucher specimen BKF no. 127614). The plant material was collected in accordance to the international guideline for plant specimen, carried out by Dr. Thawatchai Santisuk of The Forest Herbarium, Royal Forest Department, Ministry of Natural Resource and Environment, Bangkok, Thailand (23). The detailed of chemical synthesis of the reported derivatives can be found in the **Supplementary Method**.

***In silico* target prediction**

Potential targets of CTA and its derivatives were predicted using the SwissTargetPrediction server (37) (<https://www.swisstargetprediction.ch>) accessed on 19 July 2024). The Simplified Molecular Input Line Entry System (SMILES) string of CTA was retrieved from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov>) and served as the query molecule, which was compared against known ligands across a comprehensive library of protein targets (37). The prediction was performed selectively to CTA based on assumption that derivatives sharing the similar core scaffold would exhibit similar target profiles. The prediction generated the top 100 most probable targets which were classified into groups or families for further computational analysis.

System preparation and molecular docking

The X-ray crystal structures of the target proteins (**Table S1 – Supplementary data 1**) were retrieved from the RCSB Protein Data Bank (<https://www.rcsb.org>). Prior to molecular docking, all solvent molecules and co-crystallized ligands were removed. The SMILES strings of CTA and its derivatives were converted into three-dimensional (3D) structures and subsequently energy-minimized using the Molecular Mechanics 2 (MM2) force field via Chem3D software. The minimized compounds and prepared target proteins were converted to PDBQT format utilizing the AutoDockFR 1.0 software suite (58). Rigid molecular docking was executed using AutoDock Vina 1.2.7 (59). The docking grid box was uniformly set to $20 \times 20 \times 20$ Å, with specific grid center coordinates (x, y, z) detailed in **Supplementary Table S1 (Supplementary data 1)**. The search exhaustiveness was set to 64, while all other algorithmic settings were retained at their defaults. Finally, the binding modes and intermolecular interactions between each compound and the target proteins were analyzed and visualized using UCSF ChimeraX (60) for 3D representations and BIOVIA Discovery Studio Visualizer (BIOVIA, San Diego, CA, USA) for 2D interaction diagrams.

Molecular dynamics (MD) simulation

MD simulations were performed under periodic boundary conditions using the PMEMD.cuda module in AMBER 24. Partial atomic charges for each ligand were calculated using the AM1-BCC method (61) via the Antechamber module. The protonation states of ionizable residues at pH 7.0 were predicted using

PROPKA through the PDB2PQR server (62). The General AMBER Force Field 2 (GAFF2) (61) and the AMBER ff19SB force field (63) were applied to the ligands and proteins, respectively. Missing hydrogen atoms were added using the LEaP module in AMBER 24. Each system was solvated in a TIP3P water box extending 10 Å from the protein surface, and chloride ions (Cl⁻) were added to neutralize the system. Following standard protocols (51, 64) the systems underwent sequential energy minimization, thermalization, and density equilibration. The production phase was then conducted under an isothermal-isobaric (NPT) ensemble at 310 K and 1 atm until a total simulation time of 200 ns was reached. MD trajectories were saved every 500 steps for subsequent analysis. Structural evaluations over the simulation period were performed using the CPPTRAJ module (65). The total binding free energies and per-residue energy decompositions of the protein-ligand complexes were computed utilizing the MMPBSA.py module (55).

Surface plasmon resonance

Binding interactions between test compounds and the recombinant epidermal growth factor receptor (EGFR) kinase domain were characterized by surface plasmon resonance (SPR) on a Biacore X100 instrument (Cytiva, USA) operated at 25 °C in phosphate-buffered saline (PBS) supplemented with 0.05% Tween-20 and 5% (v/v) dimethyl sulfoxide (DMSO). The buffer was filtered through a 0.22 µm membrane and degassed prior to use. the same buffer composition was used as the analyte matrix throughout to keep the bulk refractive index matched between sample and reference. His-tagged recombinant EGFR (MedChemExpress, USA) was immobilized on sensor Chip NTA (Cytiva, USA). Test compounds were supplied as 10 mM stock solutions in 100% DMSO and stored at -20 °C.

The sensor Chip NTA was conditioned with 350 mM EDTA in water, injected over the active (Fc2) and reference (Fc1) flow cells for 60 seconds at 30 µL/min. The chip was loaded with 0.5 mM NiCl₂ in running buffer over Fc2 only for 60 seconds at 5 µL/min, followed by another wash with 3 mM EDTA in running buffer to remove non-chelated nickel. The compounds were injected over the nickel-only Fc2 surface (120 s association / 120 s dissociation, 30 µL/min) to confirm that no measurable binding occurred to the nickel-NTA matrix in the absence of ligand. Initial binding screenings were performed using association time at 120 s, dissociation time 120 s, flow rate at 30 µL/min, and analyte concentrations at 1 µM and 10 µM, in order to identify specific binders above the buffer-blank baseline and demonstrated dose-dependent responses.

For affinity measurements, a five-point serial titration at 25, 50, 75, 125 and 250 μM , was then acquired for the confirmed binders from the initial screen. Sensorgrams were processed in Biacore X100 Evaluation Software (v2.0.x) with Fc1 reference subtraction, buffer-blank ("double") referencing. Dissociation constant was obtained from a steady-state single-site affinity analysis.

Funding

This research project has been supported by Mahidol University to P.S. (Fundamental Fund: fiscal year 2025; Grant No. FF68-091) by National Science Research and Innovation Fund (NSRF). The project received partial funding from the National Research Council of Thailand (NRCT) to S.C. (Grant No. N35A650038) and received supports from the Reinventing University Project via the Faculty of Science, Mahidol University and from the Ramathibodi Foundation.

Author contributions

BP: conduct *in silico* analysis, writing—original draft preparation

DL: synthesis and characterization of diphyllin derivatives

NC: Isolation of cleistanthin A, synthesis and characterization of cleistanthin A derivatives

SK: conduct and analysis of surface plasmon resonance (SPR) experiment, interpretation of data

BM: Isolation of cleistanthin A, synthesis and characterization of cleistanthin A derivatives

PT: supervision of synthesis and characterization of cleistanthin A derivatives

SC: compound design and supervision of diphyllin derivatives preparation

NPM: conceptualization, interpretation of data

SM: conceptualization, interpretation of data

SR: conceptualization, interpretation of data

PS: conceptualization, interpretation of data, writing - review and editing, funding acquisition.

BN: conceptualization, supervision of *in silico* analysis, interpretation of data writing - review and editing

All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Data Availability

All datasets and parameters supporting the conclusions of this study are publicly available at <https://github.com/bunnatutpan/EGFR-as-a-Potential-Molecular-Target-of-Cleistanthin-A-and-Diphyllin-Derivatives.git>

Abbreviations

1.05X	5% more concentrated than standard working buffer
1X	Normal concentration of standard working buffer
Å	Angstrom
A ... D	Acceptor-to-donor distance
A ... H–D	Acceptor-hydrogen-donor angle
ALK	Anaplastic lymphoma kinase
ATP	Adenosine triphosphate
BRAF	v-Raf murine sarcoma viral oncogene homolog B
°C	Degrees Celsius
CADD	Computer-aided drug design
C797S	Cysteine 797 to Serine substitution mutation

Cl^-	Chloride ion
CTA	Cleistanthin A
CDCl_3	Deuterated chloroform
Cs_2CO_3	Cesium carbonate
δ	Chemical shifts
Del19 (EGFR)	Exon 19 deletion
DMF	Dimethylformamide
DIPEA	<i>N,N'</i> -diisopropylethylamine
DMAP	Dimethylaminopyridine
DMSO	Dimethyl sulfoxide
EDC/NHS	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and N-hydroxysuccinimide
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EtOAc	Ethyl acetate
ΔE_{ele}	Electrostatic interactions
ΔE_{vdW}	Van der Waals interactions
FAK	Focal adhesion kinase
Fc	Flow cell
GAFF2	General AMBER Force Field 2
GB / GBSA	Generalized Born / Generalized Born Surface Area
ΔG_{bind}	Binding free energy
ΔE_{ele}	Polar electrostatic interactions
$\Delta G_{\text{sol}}^{\text{ele}}$	Polar solvation effect
ΔE_{MM}	Gas-phase interaction energy
$G_{\text{sol}}^{\text{nonpolar}}$	Non-polar solvation effect
$\Delta G_{\text{sol}}^{\text{ele}}$	Polar solvation term
$\Delta G_{\text{sol}}^{\text{nonpolar}}$	Non-polar solvation effect
ΔG_{sol}	Total solvation free energy
ΔG_{total}	Total binding free energies

ΔE_{vdW}	Van der Waals interactions
kcal/mol	Kilocalories per mole
K_D	Dissociation constant
KI	Potassium iodide
kJ/mol	Kilojoules per mole
K_2CO_3	Potassium carbonate
L858R	Leucine 858 to Arginine substitution mutation
MD	Molecular dynamics simulation
cMET	Hepatocyte growth factor receptor
MM	Molecular Mechanics
MM/PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MM/GBSA	Molecular Mechanics generalized Born Surface Area
NaCl	Sodium chloride
Ni-NTA	Nickle-Nitrilotriacetic Acid
NiCl ₂	Nickel(II) chloride
NMR	Nuclear magnetic resonance
Ns	Nanosecond
NSCLC	Non-small cell lung cancer
PBS	Phosphate-buffered saline
PBS-P+	Phosphate-buffered saline with 0.05% Surfactant P20 and 0.5 mg/mL
PDB	Protein Data Bank
PDB2PQR	Structure-preparation server (protonation states)
PDBQT	PDB with partial charges and atom types (AutoDock format)
RET	Rearranged during transfection
R_g	Radius of gyration
RMSD	Root-Mean-Square Deviation
ROS1	ROS proto-oncogene 1, receptor tyrosine kinase
RU	Resonance Unit
SMILES	Simplified Molecular Input Line Entry System
SPR	Surface plasmon resonance
TLC	Thin-layer chromatography

TMS	Tetramethylsilane
T790M	Threonine 790 to Methionine substitution mutation
vATPase	Vacuolar-type H ⁺ -ATPase

Reference

1. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*. 2022;12(7):3049-62.
2. Sertkaya A, Beleche T, Jessup A, Sommers BD. Costs of Drug Development and Research and Development Intensity in the US, 2000-2018. *JAMA Netw Open*. 2024;7(6):e2415445.
3. Lakhan SE. In Silico Research Is Rewriting the Rules of Drug Development: Is It the End of Human Trials? *Cureus*. 2025;17(5):e84007.
4. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational Methods in Drug Discovery. *Pharmacological Reviews*. 2014;66(1):334-95.
5. Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today*. 2017;22(2):249-69.
6. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature*. 2023;616(7958):673-85.
7. Hanahan D. Hallmarks of cancer—Then and now, and beyond. *Cell*. 2026;189(8):2254-77.
8. Sylva RM, Ahmed SI. Targeted therapies in cancer. *Surgery (Oxford)*. 2024;42(3):150-5.
9. Gross S, Rahal R, Stransky N, Lengauer C, Hoeflich KP. Targeting cancer with kinase inhibitors. *The Journal of Clinical Investigation*. 2015;125(5):1780-9.
10. Nicolini A, Ferrari P, Silvestri R, Bini DA. Targeted Therapies in the Most Common Advanced Solid Tumors, Drug Resistance, and Counteracting Progressive Micrometastatic Disease: The Next Frontier of Research. *MedComm (2020)*. 2025;6(10):e70373.
11. Kumar A, Kumar A. Non-small-cell lung cancer-associated gene mutations and inhibitors. *Advances in Cancer Biology - Metastasis*. 2022;6:100076.
12. Waarts MR, Stonestrom AJ, Park YC, Levine RL. Targeting mutations in cancer. *J Clin Invest*. 2022;132(8).

13. Ahmad M, Tahir M, Hong Z, Zia MA, Rafeeq H, Ahmad MS, et al. Plant and marine-derived natural products: sustainable pathways for future drug discovery and therapeutic development. *Front Pharmacol*. 2024;15:1497668.
14. Wali AFF, Talath S, Babiker R, El-Tanani M, Rangraze IR, Ibraheem W, et al. Natural products as kinase inhibitors in lung cancer: molecular mechanisms, therapeutic potential, and clinical trials. *Frontiers in Pharmacology*. 2026;Volume 17 - 2026.
15. Baier A, Szyszka R. Compounds from Natural Sources as Protein Kinase Inhibitors. *Biomolecules*. 2020;10(11).
16. Idakwoji PA, Tukur U, Onwukwe BO, Agboola JB, Onugwu EO, Iyiola AT, et al. Phytochemicals as promising lead compounds for drug discovery: A comprehensive review of mechanisms and pharmacological potentials. *Letters in Drug Design & Discovery*. 2026;23(1):100363.
17. Ding Y, Xue X. Medicinal Chemistry Strategies for the Modification of Bioactive Natural Products. *Molecules*. 2024;29(3):689.
18. Wilcken R, Zimmermann MO, Lange A, Joerger AC, Boeckler FM. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *Journal of Medicinal Chemistry*. 2013;56(4):1363-88.
19. Dipa CD, Hossain S, Chy MMK, Rahman MSF, Kayes T, Easha AM, et al. In silico exploration of anticancer plant phytochemicals for EGFR-targeted lung cancer therapy. *Scientific Reports*. 2025;15(1):27809.
20. Kumar A, Patel SK, Karthikeyan C, Tiwari A, Moorthy NSHN. Investigation of plant-derived compounds as potent inhibitors of casein kinase-1 alpha through computational analysis and molecular dynamic simulations. *In Silico Research in Biomedicine*. 2026;2:100285.
21. Alibakhshi A, Malekzadeh R, Hosseini SA, Yaghoobi H. Investigation of the therapeutic role of native plant compounds against colorectal cancer based on system biology and virtual screening. *Scientific Reports*. 2023;13(1):11451.
22. Pradheepkumar CP, Shanmugam G. Anticancer potential of cleistanthin A isolated from the tropical plant *Cleistanthus collinus*. *Oncol Res*. 1999;11(5):225-32.
23. Tuchinda P, Kumkao A, Pohmakotr M, Sophasan S, Santisuk T, Reutrakul V. Cytotoxic aryl naphthalide lignan glycosides from the aerial parts of *Phyllanthus taxodiifolius*. *Planta Med*. 2006;72(1):60-2.
24. Jearawuttanakul K, Khumkhong P, Suksen K, Reabroi S, Munyoo B, Tuchinda P, et al. Cleistanthin A induces apoptosis and suppresses motility of colorectal cancer cells. *Eur J Pharmacol*. 2020;889:173604.
25. Liu S, Wang L, Ding W, Wang D, Wang X, Luo Q, et al. Cleistanthin A inhibits the invasion of MDA-MB-231 human breast cancer cells: involvement of the β -catenin pathway. *Pharmacological Reports*. 2020;72(1):188-98.
26. Lourthuraj AA, Selvam MM, Ravikrishnan B. Analyses of Antiproliferative Property of Cleistanthin-A Against A549 Cell Line. *Biomedical and Pharmacology Journal*. 2017;10:255-9.
27. Pradheepkumar CP, Panneerselvam N, Shanmugam G. Cleistanthin A causes DNA strand breaks and induces apoptosis in cultured cells. *Mutat Res*. 2000;464(2):185-93.
28. Hou W, Huang LJ, Huang H, Liu SL, Dai W, Li ZM, et al. Bioactivities and Mechanisms of Action of Diphyllin and Its Derivatives: A Comprehensive Systematic Review. *Molecules*. 2023;28(23).
29. Zhao Y, Lu Y, Ma J, Zhu L. Synthesis and Evaluation of Cleistanthin A Derivatives as Potent Vacuolar H⁺-ATPase Inhibitors. *Chemical Biology & Drug Design*. 2015;86(4):691-6.
30. Khongpraphan S, Sanongkiet S, Luangjindarat C, Thairat S, Munyoo B, Chabang N, et al. ECDD-S16, a synthetic derivative of cleistanthin A, suppresses pyroptosis in *Burkholderia pseudomallei*-infected U937 macrophages. *PLoS One*. 2025;20(7):e0327457.
31. Ekchariyawat P, Saengfak R, Sanongkiet S, Charoenwongpaiboon T, Khongpraphan S, Mala S, et al. ECDD-S16 targets vacuolar ATPase: A potential inhibitor compound for pyroptosis-induced inflammation. *PLoS One*. 2023;18(11):e0292340.

32. Paha J, Kanjanasirirat P, Munyoo B, Tuchinda P, Suvannang N, Nantasenamat C, et al. A novel potent autophagy inhibitor ECDD-S27 targets vacuolar ATPase and inhibits cancer cell survival. *Sci Rep*. 2019;9(1):9177.
33. Lindstrom A, Anantpadma M, Baker L, Raghavendra NM, Davey R, Davisson VJ. Phenotypic Prioritization of Diphyllin Derivatives That Block Filoviral Cell Entry by Vacuolar (H⁺) -ATPase Inhibition. *ChemMedChem*. 2018;13(24):2664-76.
34. Kornev AP, Taylor SS. Defining the conserved internal architecture of a protein kinase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2010;1804(3):440-4.
35. Cicenias J, Zalyte E, Bairoch A, Gaudet P. Kinases and Cancer. *Cancers (Basel)*. 2018;10(3).
36. Tomuleasa C, Tigu A-B, Munteanu R, Moldovan C-S, Kegyes D, Onaciu A, et al. Therapeutic advances of targeting receptor tyrosine kinases in cancer. *Signal Transduction and Targeted Therapy*. 2024;9(1):201.
37. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Research*. 2019;47(W1):W357-W64.
38. Friedlaender A, Perol M, Banna GL, Parikh K, Addeo A. Oncogenic alterations in advanced NSCLC: a molecular super-highway. *Biomarker Research*. 2024;12(1):24.
39. Zhao Z, Xie L, Bourne PE. Structural Insights into Characterizing Binding Sites in Epidermal Growth Factor Receptor Kinase Mutants. *Journal of Chemical Information and Modeling*. 2019;59(1):453-62.
40. Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, et al. A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib). *Cancer Research*. 2004;64(18):6652-9.
41. Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MF, et al. Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain. *PLoS Medicine*. 2005;2(3):e73.
42. Emami L, Khabnadideh S, Faghih Z, Farahvasi F, Zonobi F, Gheshlaghi SZ, et al. Synthesis, biological evaluation, and computational studies of some novel quinazoline derivatives as anticancer agents. *BMC Chemistry*. 2022;16(1).
43. Todsaporn D, Zubenko A, Kartsev VG, Mahalapbutr P, Geronikaki A, Sirakanyan SN, et al. Furopyridine Derivatives as Potent Inhibitors of the Wild Type, L858R/T790M, and L858R/T790M/C797S EGFR. *The Journal of Physical Chemistry B*. 2024;128(50):12389-402.
44. Hou T, Wang J, Li Y, Wang W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*. 2011;51(1):69-82.
45. Todsaporn D, Mahalapbutr P, Poo-Arporn RP, Choowongkomon K, Rungrotmongkol T. Structural dynamics and kinase inhibitory activity of three generations of tyrosine kinase inhibitors against wild-type, L858R/T790M, and L858R/T790M/C797S forms of EGFR. *Computers in Biology and Medicine*. 2022;147:105787.
46. Suriya U, Mahalapbutr P, Wimonson W, Yotphan S, Choowongkomon K, Rungrotmongkol T. Quinoxalinones as A Novel Inhibitor Scaffold for EGFR (L858R/T790M/C797S) Tyrosine Kinase: Molecular Docking, Biological Evaluations, and Computational Insights. *Molecules*. 2022;27(24):8901.
47. Nili Ahmadabadi M, Rezaee E, Nematpour M, Karami L, Mokhtari S, Kobarfard F, et al. Synthesis, Molecular Dynamics Simulation, and In-vitro Antitumor Activity of Quinazoline-2,4,6-triamine Derivatives as Novel EGFR Tyrosine Kinase Inhibitors. *Iranian Journal of Pharmaceutical Research*. 2023;21(1).
48. Mahalapbutr P, Leechaisit R, Thongnum A, Todsaporn D, Prachayasittikul V, Rungrotmongkol T, et al. Discovery of Anilino-1,4-naphthoquinones as Potent EGFR Tyrosine Kinase Inhibitors: Synthesis, Biological Evaluation, and Comprehensive Molecular Modeling. *ACS Omega*. 2022;7(21):17881-93.

49. Todsaporn D, Zubenko A, Kartsev V, Aiebchun T, Mahalapbutr P, Petrou A, et al. Discovery of Novel EGFR Inhibitor Targeting Wild-Type and Mutant Forms of EGFR: In Silico and In Vitro Study. *Molecules*. 2023;28(7):3014.
50. Shi K, Wang G, Pei J, Zhang J, Wang J, Ouyang L, et al. Emerging strategies to overcome resistance to third-generation EGFR inhibitors. *Journal of Hematology & Oncology*. 2022;15(1).
51. Nutho B, Wilasluck P, Deetanya P, Wangkanont K, Arsakhant P, Saeeng R, et al. Discovery of C-12 dithiocarbamate andrographolide analogues as inhibitors of SARS-CoV-2 main protease: *in vitro* and *in silico* studies. *Computational and Structural Biotechnology Journal*. 2022;20:2784-97.
52. Wongpan A, Panvongsa W, Krobthong S, Nutho B, Kanjanasirirat P, Jearawuttanakul K, et al. Cleistanthin A derivative disrupts autophagy and suppresses head and neck squamous cell carcinoma progression via targeted vacuolar ATPase. *Sci Rep*. 2024;14(1):22582.
53. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*. 2015;10(5):449-61.
54. Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem*. 2004;25(2):238-50.
55. Miller BR, III, McGee TD, Jr., Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation*. 2012;8(9):3314-21.
56. Günther M, Lategahn J, Juchum M, Döring E, Keul M, Engel J, et al. Trisubstituted Pyridinylimidazoles as Potent Inhibitors of the Clinically Resistant L858R/T790M/C797S EGFR Mutant: Targeting of Both Hydrophobic Regions and the Phosphate Binding Site. *Journal of Medicinal Chemistry*. 2017;60(13):5613-37.
57. Wang S, Tsui ST, Liu C, Song Y, Liu D. EGFR C797S mutation mediates resistance to third-generation inhibitors in T790M-positive non-small cell lung cancer. *Journal of Hematology & Oncology*. 2016;9(1):59.
58. Ravindranath PA, Forli S, Goodsell DS, Olson AJ, Sanner MF. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Computational Biology*. 2015;11(12):e1004586.
59. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*. 2021;61(8):3891-8.
60. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science*. 2023;32(11):e4792.
61. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*. 2002;23(16):1623-41.
62. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*. 2007;35(Web Server):W522-W5.
63. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migués AN, et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation*. 2020;16(1):528-52.
64. Nutho B, Tungmunthum D. Exploring Major Flavonoid Phytochemicals from *Nelumbo nucifera* Gaertn. as Potential Skin Anti-Aging Agents: In Silico and In Vitro Evaluations. *International Journal of Molecular Sciences*. 2023;24(23):16571.
65. Roe DR, Cheatham TE, III. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*. 2013;9(7):3084-95.

Supplementary Files

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

- [SwissTargetPredictionCTA.csv](#)
- [SupplementaryMethods.docx](#)
- [SupplementaryData2.docx](#)
- [SupplementaryData1.docx](#)