

Article

Not peer-reviewed version

Exploiting the T790M Mutation: Sulfur-Anchoring as a New Paradigm for Non-Covalent EGFR Inhibition

[Shrikant S. Nilewar](#) , [Shuvadip Khanra](#) , [Manav Pandya](#) , [Sandesh Lodha](#) , [Perli Kranti Kumar](#) , [Nagaraju Bandaru](#) , [Antonio Jose Naranjo-Redondo](#) , [Ricardo Pérez-Pastén-Borja](#) * , [Tushar Janardan Pawar](#) *

Posted Date: 29 April 2026

doi: 10.20944/preprints202604.2046.v1

Keywords: EGFR T790M; molecular dynamics; generative artificial intelligence; sulfur anchoring; drug resistance; free energy landscape; non-covalent inhibitors



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Exploiting the T790M Mutation: Sulfur-Anchoring as a New Paradigm for Non-Covalent EGFR Inhibition

Shrikant S. Nilewar¹, Shuvadip Khanra¹, Manav Pandya¹, Sandesh Lodha¹,
Perli Kranti Kumar², Nagaraju Bandaru³, Antonio Jose Naranjo-Redondo⁴,
Ricardo Pérez-Pastén-Borja^{5,*} and Tushar Janardan Pawar^{4,6,*}

¹ Department of Pharmaceutical Chemistry, Maliba Pharmacy College, Uka Tarsadia University, Bardoli 394350, Gujarat, India

² Department of Pharmaceutical Analysis, J.K.K. Nattraja College of Pharmacy, Kumarapalayam, 638183, Tamil Nadu, India

³ Department of Pharmacology, Sree Dattha Institute of Pharmacy Sheriguda, Ibrahimpatnam, Hyderabad, 501510, Telangana, India

⁴ División de Ingeniería, Universidad Anáhuac Querétaro, Circuito Universidades I, Fracción 2 S/N, Zibatá, El Marqués 76246, Querétaro, Mexico

⁵ Departamento de Farmacia, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Unidad Académica Adolfo López Mateos, Del. GAM, Mexico City, 77380 Mexico

⁶ Centro de Investigación, Universidad Anáhuac Querétaro, Circuito Universidades I, Fracción 2 S/N, Zibatá, El Marqués 76246, Querétaro, Mexico

* Correspondence: rpastenb@ipn.mx (R.P.-P.-B); tushar.pawar@anahuac.mx (T.J.P.).

Abstract

Background/Objectives: The EGFR T790M gatekeeper mutation drives resistance in non-small cell lung cancer by sterically hindering inhibitors and restoring ATP affinity. As emerging C797S mutations render covalent inhibitors obsolete, novel non-covalent strategies are critical. This study identifies inhibitors that redefine the mutant methionine sulfur atom as a primary stabilizing anchor rather than a liability. **Methods:** A reinforcement learning-based generative AI framework (DrugEx) sampled 100,000 molecules, prioritized through QSAR-based classification (mean ROC-AUC: 0.91 ± 0.01) and physicochemical filtering. Four lead candidates underwent 200 ns all-atom molecular dynamics (MD) simulations. Mechanistic stability and energetic convergence were quantified via Free Energy Landscape (FEL) analysis, post-simulation Ramachandran validation, and ensemble-averaged Molecular Mechanics Generalized Born Surface Area (MM-GBSA) binding free energy calculations. **Results:** Candidate **106** demonstrated high mutation tolerance by redistributing interactions toward the Met790 sulfur atom, contrasting with reference scaffolds **14** and **88** which suffered significant affinity loss. MD analysis revealed that potency is fundamentally dictated by successful recruitment of the thioether environment, locking the complex within a narrow, well-defined thermodynamic basin. Candidate **106** maintained stable binding (-11.0 kcal/mol by docking) corroborated by a robust MM-GBSA ΔG_{bind} of -50.51 kcal/mol, primarily driven by persistent π -sulfur contacts (85% occupancy). Ramachandran analysis confirmed over 90% favored region occupancy, ensuring interactions occur without non-physical protein strain. **Conclusions:** Successful T790M/C797S resistance bypass is achievable by exploiting the gatekeeper methionine's electronic environment. This mutation-aware blueprint provides a non-covalent strategy that avoids the metabolic and resistance-prone liabilities of covalent warheads.

Keywords: EGFR T790M; molecular dynamics; generative artificial intelligence; sulfur anchoring; drug resistance; free energy landscape; non-covalent inhibitors

1. Introduction

The clinical management of non-small cell lung cancer (NSCLC) has been fundamentally redefined by the discovery of oncogenic drivers within the kinase domain of the epidermal growth factor receptor (EGFR) [1,2]. The advent of tyrosine kinase inhibitors (TKIs) targeting these sensitizing mutations, such as exon 19 deletions and the L858R point mutation, initially provided a paradigm shift in patient survival, offering robust and selective therapeutic responses [3,4]. These small molecules function by competing with adenosine triphosphate (ATP) for binding within the catalytic cleft, thereby preventing the phosphorylation of downstream signaling proteins that drive uncontrolled cellular proliferation [5]. However, the clinical utility of these agents is perennially truncated by the inevitable emergence of acquired resistance, which remains the most formidable obstacle in precision oncology [6,7].

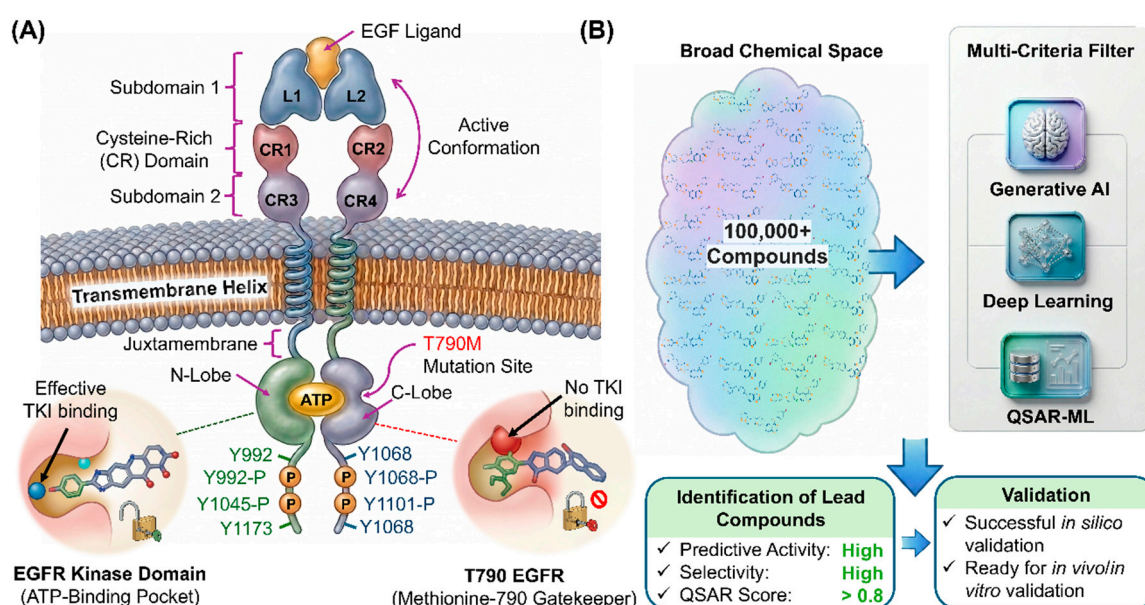


Figure 1. Translational landscape and mechanistic basis for mutation-aware EGFR inhibition. (A) Represents the progression of clinical resistance in NSCLC from sensitizing mutations to the T790M gatekeeper variant. (B) Details the discovery workflow for lead compounds.

Among the various resistance pathways, the T790M gatekeeper mutation is the most prevalent, occurring in approximately 50% to 60% of patients following treatment with first- and second-generation TKIs [8,9]. This specific substitution replaces a polar threonine residue with a bulky, hydrophobic methionine at the entry to the ATP-binding cleft (Figure 1A). The impact of this mutation is twofold: it introduces significant steric hindrance that prevents the binding of classic inhibitors and, more critically, it restores the receptor's affinity for ATP to wild-type levels, effectively outcompeting non-covalent inhibitors (Figure 1A) [10]. The methionine side chain effectively acts as a gatekeeper that blocks the deep hydrophobic pocket, a region previously accessible to earlier inhibitors [10,11].

To combat this, third-generation covalent inhibitors, such as osimertinib, were engineered to form a stable thioether bond with the Cys797 residue located at the edge of the binding pocket [12,13]. While these agents initially achieved high objective response rates, the rapid clinical ascent of the C797S mutation [14], which replaces the nucleophilic cysteine with a serine has rendered even these last-line therapies obsolete for many late-stage patients [15,16]. Because serine lacks the thiol group necessary for covalent attachment to acrylamide-based warheads, the inhibitors lose their primary anchoring mechanism, leading to a dramatic reduction in potency [17].

This recurring cycle of resistance underscores a critical need for a departure from the covalent-only design philosophy [18,19]. The pursuit of high-affinity, non-covalent inhibitors capable of bypassing both T790M and C797S resistance requires a sophisticated re-evaluation of the altered physiochemical environment within the mutant ATP-binding cleft [20]. Traditional drug design has historically treated the T790M mutation as a liability, a bump to be avoided through steric maneuvering. However, this perspective overlooks the unique electronic opportunities presented by the methionine side chain. Unlike the native threonine, methionine introduces an electron-rich thioether sulfur atom that significantly modulates the electrostatic potential and London dispersion forces of the gatekeeper region [21–23]. Establishing stable interactions with the sulfur atom of Met790 can serve as a potent anchoring mechanism, potentially compensating for the loss of the native threonine-mediated hydrogen bond network (Figure 1A).

The methionine sulfur atom is a unique atomic chameleon in molecular recognition. It can participate in π -sulfur interactions, where the electron-rich sulfur interacts with the aromatic rings of an inhibitor, or in specialized hydrophobic contacts that are more robust than typical carbon-carbon interactions [24,25]. Exploiting these interactions requires an inhibitor scaffold that is not only sterically compatible with the restricted gatekeeper geometry but also electronically tuned to favor these specialized non-covalent bonds. This represents a transition from mutation-avoidance to mutation-exploitation, a strategy that could yield inhibitors with sustained efficacy even as the receptor evolves [26].

The necessity for this type of research is driven by the limits of traditional trial-and-error medicinal chemistry. The chemical space available for kinase inhibition, estimated at 10^{60} molecules, is too vast for manual exploration [27,28]. Furthermore, the time-dependent stability of non-covalent interactions cannot be captured through static modeling or simple docking scores. The integration of generative artificial intelligence and high-resolution molecular dynamics (MD) simulations has recently emerged as a transformative necessity in drug discovery. Models such as DrugEx facilitate the design of novel scaffolds that satisfy multi-objective criteria, including drug-likeness and synthetic accessibility, while sampling regions of chemical space that have been historically under-explored [29].

By utilizing deep learning models to sample chemical space and physics-based simulations to evaluate the thermodynamic fitness of candidates, researchers can decipher the intricate time-dependent stability of interactions that define long-term clinical efficacy. These computational workflows allow for the characterization of the Free Energy Landscape (FEL), identifying the most stable conformational states of a protein-ligand complex. Without these advanced frameworks, the subtle stereochemical adaptations required to turn a resistance-inducing mutation into a therapeutic anchor remain undiscovered, and the role of ligand tautomerism remains a hidden variable in drug failure [30,31].

In this work, we present a systematic identification of novel non-covalent EGFR inhibitors specifically engineered to leverage tautomer-specific sulfur anchoring. By integrating a reinforcement learning-based generative framework (DrugEx) with a multi-stage thermodynamic and stereochemical validation pipeline, we explore a novelty-constrained chemical space to prioritize scaffolds optimized for the T790M gatekeeper (Figure 1B). We utilize extensive all-atom molecular dynamics simulations, encompassing up to 200 ns per trajectory, and FEL analysis to investigate the temporal stability and energetic convergence of these candidates. This study focuses on characterizing the structural adaptation mechanisms that allow for high-affinity binding across both wild-type and mutant variants, employing post-simulation Ramachandran analysis to ensure that these novel interactions occur within a biologically relevant and stable protein framework. By redefining the gatekeeper mutation as a chemical opportunity, our work provides a robust blueprint for the next generation of mutation-aware lung cancer therapeutics.

2. Materials and Methods

2.1. Dataset Preparation and QSAR Model Construction

Bioactivity data for the epidermal growth factor receptor (EGFR) were systematically retrieved from the Papyrus database (v5.6), which aggregates experimentally validated activity values from multiple public repositories [32]. Compounds were filtered using the corresponding UniProt identifier to ensure target specificity. To maintain data integrity, only entries featuring experimentally determined activity values and consistent assay annotations were retained for the study. Bioactivity was standardized using the mean pChEMBL values as reported by the source [33].

The data curation workflow, executed within the RDKit framework, involved the removal of duplicate molecular structures, compounds with missing activity values, and records with inconsistent annotations [34]. The finalized dataset comprised 7,555 unique compounds. For the purpose of classification modeling, compounds were categorized into active and inactive classes based on a pChEMBL threshold of 6.8, corresponding to submicromolar inhibitory potency.

Molecular descriptors were generated using circular Morgan fingerprints with a radius of two and a bit length of 2,048, selected for their proven efficacy in kinase-focused ligand modeling [35]. A Random Forest classifier was subsequently trained to discriminate between active and inactive chemotypes [36]. Model robustness was rigorously evaluated through repeated cross-validation and an independent test set. To assess the model's capacity for scaffold hopping and generalization beyond known chemical series, an additional scaffold-based validation was conducted using Bemis-Murcko frameworks [37]. In this splitting scheme, compounds sharing the same core scaffold were strictly confined to either the training or test set to prevent structural data leakage during evaluation [38].

2.2. Molecular Generation and Hierarchical Chemical Space Refinement

The validated Random Forest classification model served as the guiding objective function for molecular generation using the reinforcement learning-based DrugEx framework [39]. The generative model was initially pre-trained on the curated EGFR ligand dataset to learn the underlying chemical grammar, followed by a reinforcement learning phase to bias the sampling toward structures with high predicted inhibitory activity. This pipeline yielded an initial chemical library of 100,000 *de novo* molecules [40].

To distill this initial library into a highly enriched, developable subset, we implemented a stringent, multi-parameter hierarchical virtual screening cascade. Initial physicochemical triage established baseline oral bioavailability and pharmacokinetic viability by restricting molecular weight (250–500 Da), lipophilicity (cLogP 1.0–5.0), predicted aqueous solubility (cLogS > -6.0 at pH 7.5), and topological polar surface area (20–140 Å²), alongside strict adherence to Lipinski hydrogen-bonding thresholds. Subsequently, scaffolds were evaluated for structural tractability; we advanced only molecules with positive empirical drug-likeness scores (>0.0) while systematically purging pan-assay interference compounds (PAINS) and chemically reactive motifs [41].

Crucially, to bridge the gap between computational discovery and *in vivo* translational viability, we prioritized rigorous predictive toxicological profiling as a primary filtering checkpoint. Any compound exhibiting latent mutagenic, tumorigenic, reproductive, or severe irritant properties was unequivocally eliminated. Finally, these candidates underwent topological and conformational optimization, enforcing specific geometric compatibilities (shape index 0.2–0.8), minimizing thermodynamic entropic penalties (flexibility < 0.6), and maximizing intracellular penetrance (relative PSA < 0.5), ultimately yielding 1,601 chemically tractable compounds.

Structural novelty was explicitly quantified by comparing these compounds against the original EGFR dataset using Morgan fingerprint-based Tanimoto similarity. To prioritize distinct chemotypes, a stringent novelty threshold was applied; compounds with a Tanimoto similarity value of 0.4 or greater were excluded, resulting in a final set of 136 structurally diverse compounds prioritized for structure-based evaluation.

2.3. Molecular Docking and Active Site Preparation

Structure-based evaluation was conducted using the high-resolution crystal structure of the human EGFR kinase domain (PDB ID: 3W32) as the primary template [42]. The protein was prepared using the Protein Preparation Wizard in AutoDock Tools (v1.5.7) to assign bond orders, add missing hydrogen atoms, and optimize the hydrogen-bonding network at physiological pH (7.4) [43]. Water molecules were removed. The T790M mutant structure was generated using Swiss-PdbViewer (v4.1.0) followed by a local energy minimization of residues within a 5 Å radius to resolve steric clashes [44].

Docking simulations were performed using the MZDock engine with the SMINA scoring function, a fork of AutoDock Vina optimized for custom scoring and high-performance screening [45]. For the wild-type receptor, the binding site was centered on the coordinates of the co-crystallized ligand W32. For the T790M mutant, the grid box was defined using the same spatial constraints to ensure comparative consistency. The 136 prioritized compounds were prepared by generating 3D conformers and assigning appropriate ionization states using MarvinSketch (v25.3.5, ChemAxon, <http://www.chemaxon.com>). Docked poses were ranked according to their predicted binding affinities (kcal/mol), and the top four candidates were selected for dynamic evaluation based on their interaction patterns with the hinge region and the gatekeeper residue.

2.4. Molecular Dynamics (MD) Simulations

To evaluate the temporal stability and mechanistic adaptation of the lead compounds, all-atom molecular dynamics (MD) simulations were conducted using the Desmond engine with the OPLS4 force field [46,47]. Each protein-ligand complex was solvated in an orthorhombic box of SPC water molecules, extending 10 Å beyond the protein surface. The systems were neutralized by adding counterions (Na⁺, Cl⁻) and a physiological salt concentration of 0.15 M NaCl was maintained. The systems underwent a multi-stage equilibration protocol consisting of initial minimization under restrained conditions followed by heating and pressurized equilibration in NVT and NPT ensembles. Production runs were executed at a constant temperature of 300 K using the Nosé-Hoover thermostat and a pressure of 1.01325 bar using the Martyna-Tobias-Klein barostat. Simulations were performed for a duration of 100 ns for all complexes. For compounds **14** (wild-type), **88** (mutant), and **106** (mutant), the trajectories were extended to 200 ns to ensure the systems reached a thermodynamic equilibrium as monitored by the root-mean-square deviation (RMSD) of the ligand.

2.5. Binding Free Energy Calculation (MMPBSA)

To provide a more rigorous quantification of ligand affinity than static docking scores, ensemble-averaged binding free energies (ΔG_{bind}) were calculated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method. Calculations were performed on the stabilized production phase of the 200 ns molecular dynamics trajectories using the Prime engine with the VSGB 2.0 solvation model and the OPLS4 force field.

The total binding free energy was decomposed into its constituent physical components according to the following equation (1).

$$\Delta G_{\text{bind}} = \Delta E_{\text{internal}} + \Delta E_{\text{vwd}} + \Delta E_{\text{elec}} + \Delta G_{\text{solv}} + \Delta G_{\text{SA}} \quad (1)$$

where ΔE_{vwd} and ΔE_{elec} represent the Van der Waals and electrostatic interaction energies, respectively, and ΔG_{solv} accounts for the polar solvation energy. To ensure statistical rigor, the values were averaged over 1,000 frames extracted from the final 100 ns of each trajectory where the system had reached thermodynamic equilibrium. This approach allowed for the mathematical partitioning of the π -sulfur anchor's energetic contribution compared to the native hydrogen-bonding network.

2.6. Free Energy Landscape (FEL) Construction

The thermodynamic stability of the EGFR–ligand complexes was delineated through FEL analysis, which maps the relative stability of conformational states [48]. The FEL was constructed by projecting the MD simulation trajectories onto two principal components: the root-mean-square deviation (RMSD), representing structural displacement, and the radius of gyration (R_g), representing the compactness of the complex. The probability distribution of these variables was converted into a Gibbs free energy surface using the equation (2).

$$G(\text{RMSD}, R_g) = -k_B T \cdot \ln P(\text{RMSD}, R_g) \quad (2)$$

where k_B is the Boltzmann constant, T is the temperature, and P is the probability density. This analysis allowed for the identification of the global energy minimum and the characterization of the energetic basins occupied by the tautomer-specific inhibitors.

2.8. Statistical Analysis and Data Validation

To ensure the reproducibility and reliability of the computational findings, all quantitative results were subjected to rigorous statistical validation. QSAR model performance was evaluated using the Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) and F1-score across five-fold cross-validation and an independent test set. For MD trajectories, equilibration was statistically confirmed when the slope of the ligand RMSD reached a plateau over a continuous 50 ns window.

The stereochemical quality of the post-simulation protein-ligand complexes were validated via Ramachandran plot analysis using the Procheck algorithm to ensure that no non-physical strain was induced during the structural adaptation to the T790M mutation. Data across multiple systems were compared using one-way analysis of variance (ANOVA) where applicable, and a p -value of < 0.05 was considered the threshold for statistical significance in interaction persistence.

3. Results

3.1. QSAR Model Validation and Generative Sampling Performance

The predictive integrity of the Random Forest classification model was rigorously assessed using a multi-faceted validation strategy to ensure robust discrimination between active and inactive EGFR inhibitors. Initial evaluation via five-fold random cross-validation yielded a mean receiver operating characteristic area under the curve (ROC-AUC) of 0.91 ± 0.01 (Figure 2A). This high degree of accuracy was consistently maintained across all folds, indicating that the model successfully captured the underlying structure-activity relationships within the Papyrus-derived dataset without significant variability.

To probe the model's capacity for scaffold hopping, a prerequisite for identifying novel chemotypes, a more stringent scaffold-based validation was performed using Bemis-Murcko frameworks (Figure 2B). Under this scheme, the model retained a robust discriminative capacity with an AUC of 0.905. The negligible performance decay relative to random cross-validation suggests that the classifier prioritizes activity-relevant molecular features over localized scaffold similarities, confirming its suitability for exploring under-sampled regions of chemical space.

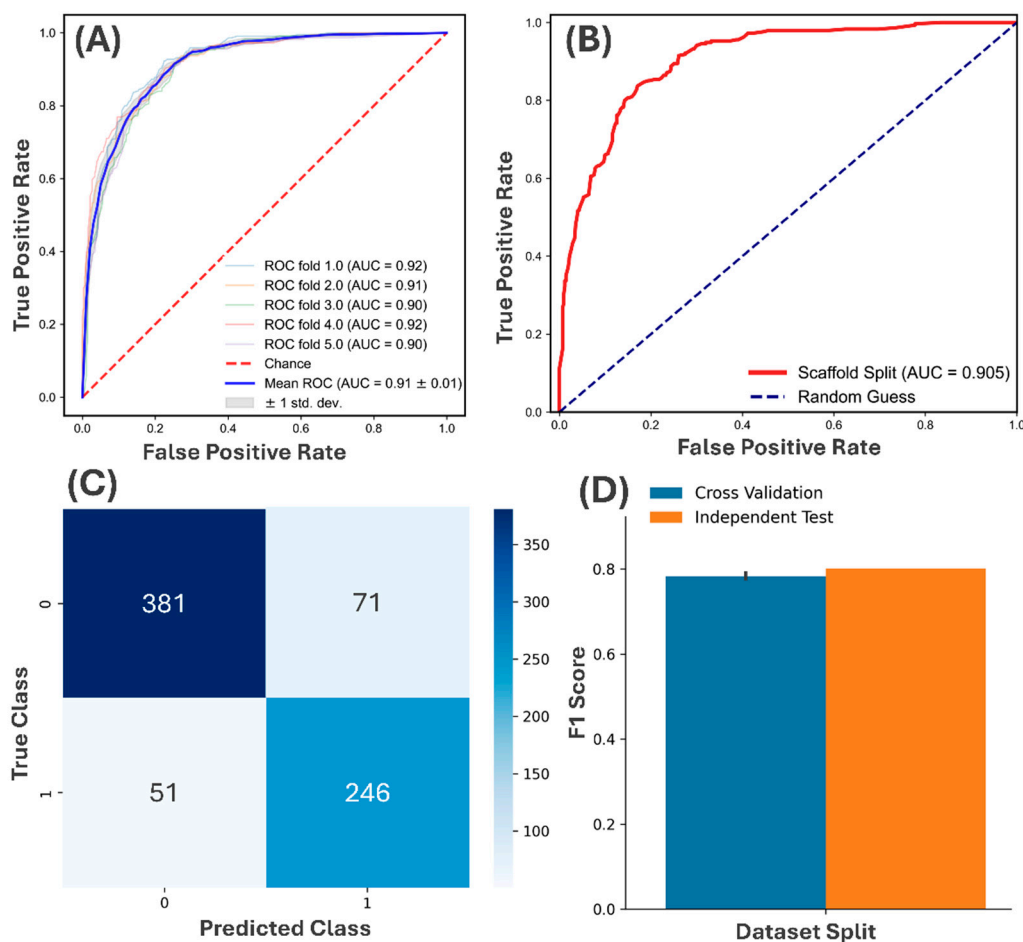


Figure 2. Statistical validation and performance metrics of the EGFR QSAR classification model. (A) Receiver Operating Characteristic (ROC) curves for five-fold random cross-validation, demonstrating a mean Area Under the Curve (AUC) of 0.91 ± 0.01 . (B) Scaffold-based ROC curve using Bemis-Murcko framework splitting. (C) Confusion matrix for the independent test set, displaying the distribution of true positives (246), true negatives (381), false positives (71), and false negatives (51), reflecting a balanced predictive accuracy. (D) Comparative analysis of F1-scores between cross-validation and the independent test set, showing consistent performance (~ 0.80) across different data partitions.

The model's reliability was further corroborated by classification metrics derived from the independent test set. The confusion matrix revealed a balanced predictive behavior, achieving 381 true negatives and 246 true positives (Figure 2C), with comparable false-positive and false-negative rates. Consistency in F1 scores between cross-validation and independent testing (approximately 0.80) further supports the model's robustness under conditions of potential class imbalance (Figure 2D).

The validated classifier was subsequently employed as the guiding function for the DrugEx reinforcement learning framework, which sampled an initial library of 100,000 candidate structures (Table 1, Stage I). This expansive chemical space was refined through a sequential filtering pipeline focused on drug-likeness and structural novelty. Physicochemical and toxicological filtering reduced the set to 1,601 compounds (Table 1, Stage II). Final assessment using a stringent Morgan fingerprint-based Tanimoto similarity threshold (< 0.4) isolated 136 compounds characterized by low similarity to previously reported EGFR inhibitors (Table 1, Stage III). These prioritized candidates combine predicted bioactivity with explicit structural novelty, providing a rationalized starting point for comparative structure-based evaluation against the T790M gatekeeper mutation.

Table 1. Summary of molecular library refinement and prioritization.

Stage	Process	Count	Rationale
I	Generative Sampling (DrugEx)	100,000	Enrichment for predicted EGFR activity.
II	Hierarchical ADMET/Liability Filtering	1,601	Optimization of solubility, polarity, and removal of toxicophores.
III	Structural Novelty Filtering	136	Exclusion of established scaffolds (Tanimoto <0.4).

3.2. Molecular Docking and Active Site Comparison

Molecular docking was employed to evaluate the binding modes and relative affinities of the 136 prioritized compounds within the ATP-binding cleft of both wild-type and T790M EGFR. All lead candidates were confirmed to occupy the canonical kinase active site, maintaining an orientation consistent with established adenine-pocket inhibitors despite significant structural divergence from the co-crystallized scaffold W32 (Figure 3A).

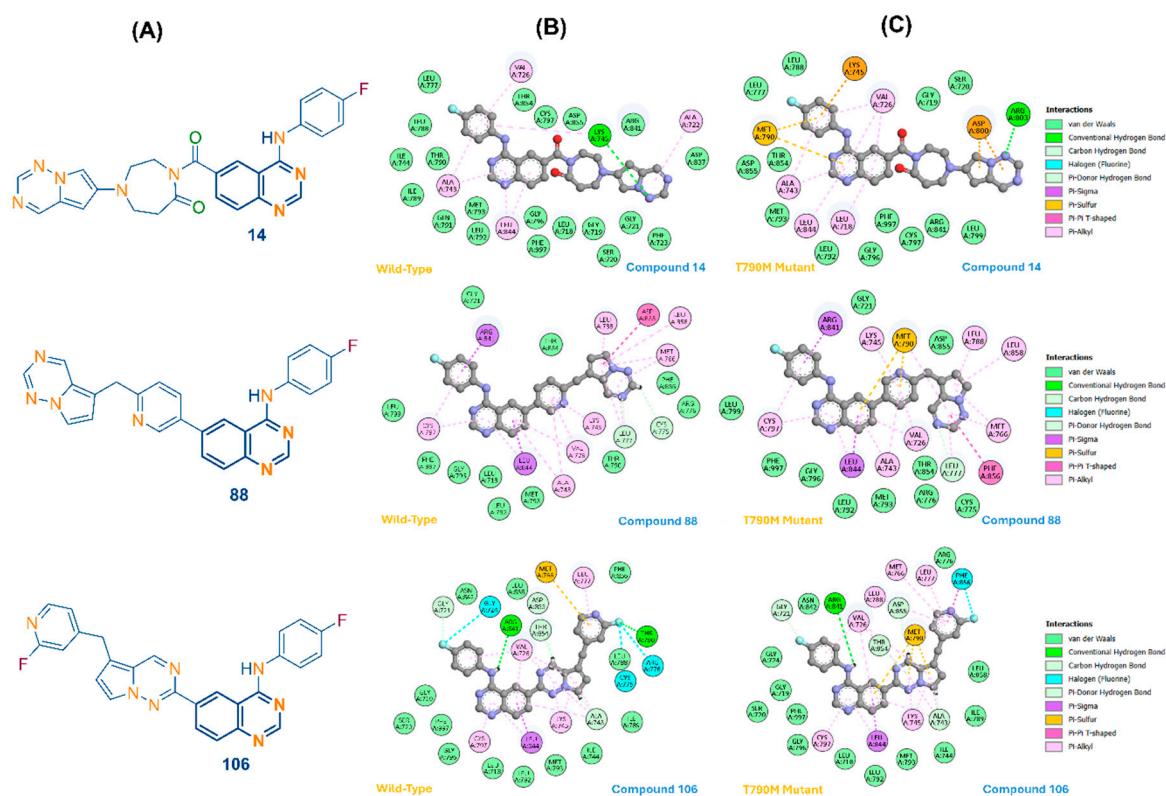


Figure 3. Comparative Binding Analysis of Lead Candidates in wild-type and T790M EGFR. (A) Chemical Structures of compounds **14**, **88**, and **106**. The 2D binding poses and molecular docking interactions of each lead within the catalytic pocket of the (B) wild-type EGFR and (C) T790M mutant EGFR.

The comparative analysis revealed distinct patterns of mutation tolerance (Figure 3B,C). Compound **14**, while exhibiting high affinity for the WT receptor (-12.2 kcal/mol), showed a significant reduction in binding strength to -9.6 kcal/mol in the T790M mutant (Table 2). This loss of potency is attributed to localized steric repulsion from the bulky Met790 side chain and the absence of compensatory interactions. Similarly, Compound **88** suffered an affinity collapse from -12.3 kcal/mol to -9.6 kcal/mol, reflecting a failure to adapt to the altered physiochemical environment of the gatekeeper region.

Table 2. Predicted binding affinities (kcal/mol) across wild-type and T790M EGFR variants.

Compound ID	WT Affinity (kcal/mol)	T790M Affinity (kcal/mol)	Δ Affinity (kcal/mol)	Primary Mutant Anchor
14	-12.2	-9.6	2.6	Lys745 (H-bond)
88	-12.3	-9.6	2.7	Lys745 (H-bond)
106	-12.3	-11.0	1.3	Met790 (π -Sulfur)
W32 (Ref)	-12.1	-10.8	1.3	Met790 (Hydrophobic)

Binding affinities were calculated using the SMINA scoring function. Δ Affinity represents the absolute difference in binding energy between wild-type (WT) and T790M systems; lower values indicate higher mutation tolerance. π -Sulfur interactions involve the electron-rich thioether of Met790 and the aromatic system of the ligand.

In contrast, Compound **106** demonstrated superior resilience, with affinity of -11.0 kcal/mol in the mutant system (Table 2). This stability is driven by an adaptive interaction redistribution. In the WT system, Compound **106** establishes a hydrogen bond with Thr790; however, upon mutation, it undergoes a structural reorganization to form a direct π -sulfur interaction with the Met790 thioether atom (Figure 3C). This interaction, supplemented by halogen bonding and extensive hydrophobic contacts with residues such as Met766 and Leu777, effectively offsets the loss of the native threonine-mediated polar network.

These docking results provide the initial structural rationale for the potency of Candidate **106**. The ability of this scaffold to recruit the methionine sulfur as a stabilizing anchor differentiates it from traditional scaffolds and provides the basis for the subsequent dynamic and thermodynamic evaluations.

3.3. Molecular Dynamics and Interaction Persistence

The structural integrity and temporal stability of the lead EGFR-ligand complexes were evaluated through all-atom molecular dynamics (MD) simulations. Root-mean-square deviation (RMSD) analysis of the protein backbones across all systems confirmed that the kinase domain remained stable throughout the trajectories, typically fluctuating between 2.0 and 3.5 Å. However, ligand-specific RMSD profiles revealed a stark contrast in mutation tolerance (Figure 4, Left). Compound **14**, despite its initial WT stability, exhibited a dramatic positional shift in the T790M mutant system, with its RMSD escalating to 8.0 Å (Figure 4A), indicating a total loss of pocket occupancy driven by steric repulsion from the methionine side chain.

In contrast, Compound **106** demonstrated remarkable resilience within the mutant catalytic cleft (Figure 4C). Compound **106** achieved a stable equilibrium after 135 ns, maintaining a ligand RMSD of 3.0 Å. This stability is corroborated by the persistent recruitment of the Met790 sulfur atom. Interaction histograms indicate that the π -sulfur contact between **106** and the gatekeeper residue remained occupied for over 85% of the 200 ns trajectory. This mechanistic adaptation effectively offsets the loss of the native Thr790 hydrogen bond, preserving the binding affinity through a transition to a robust hydrophobic and sulfur-based anchoring network.

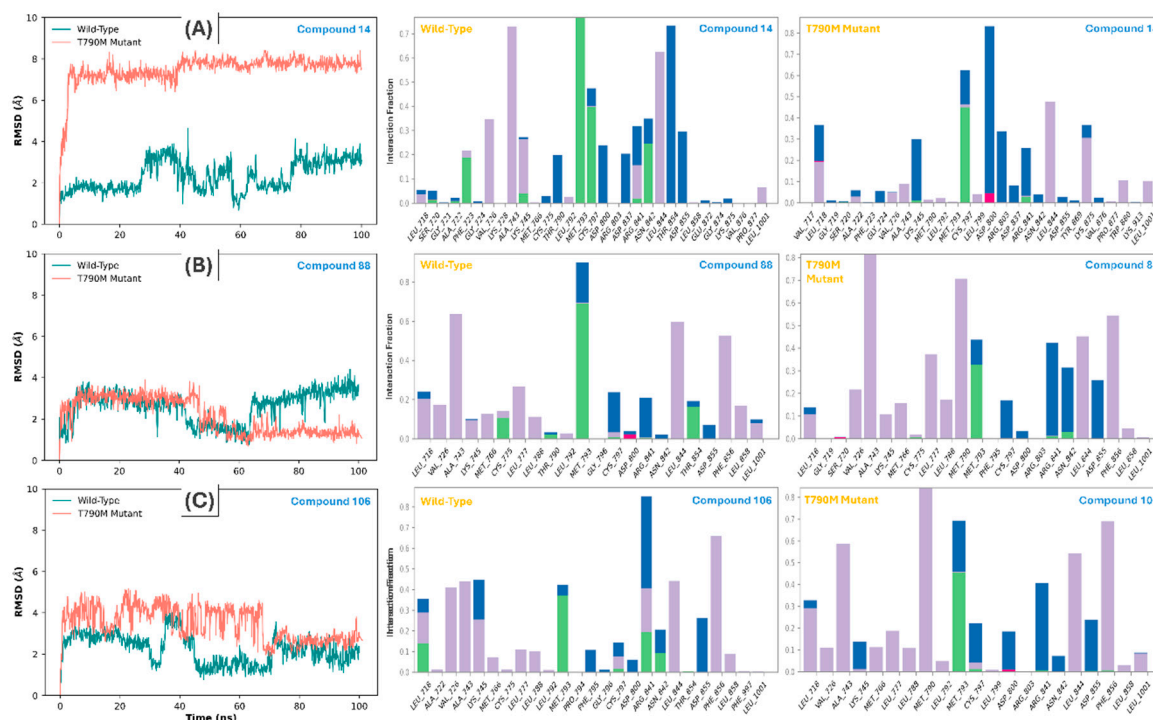


Figure 4. Molecular dynamics trajectories and protein–ligand contact profiling. (Left) Root Mean Square Deviation (RMSD) plots for Compounds (A) **14**, (B) **88**, and (C) **106** over a 100 ns simulation period, comparing the wild-type (Teal) and T790M mutant (Coral) EGFR systems. (Middle and Right) Corresponding box-and-whisker plots representing the distribution and frequency of total protein–ligand contacts throughout the production phase of the trajectory. The horizontal lines within the boxes denote the median contact number, while the whiskers indicate the 1.5 \times interquartile range (IQR). All data are presented with a shared Y-axis for the RMSD panels to facilitate direct magnitude comparison across the chemical series.

The comparative analysis between scaffolds **14**, **88**, and **106** highlights the decisive role of structural architecture in mutation bypass (Figure 4, middle and right). While Compound **88** required a significant conformational adjustment (115 ns) to achieve moderate stability, Compound **106** maintained a rigid, buried binding mode throughout the simulation. The persistence of Met790 and Met793 contacts throughout the **106** trajectory confirms that this specific scaffold is optimally suited to utilize the methionine side chain as a stabilizing anchor.

3.4. Thermodynamic Basins and Stereochemical Quality

The thermodynamic landscape of the EGFR–ligand ensembles was delineated via Free Energy Landscape (FEL) analysis, mapping the Gibbs free energy (ΔG) as a function of the RMSD and radius of gyration (R_g). This approach provides a rigorous quantitative basis for the structural transitions observed during the 200 ns production runs. For the wild-type systems, lead candidates occupied deep, singular global energy minima, where $\Delta G \approx 0$ kcal/mol represents the normalized relative free energy of the most probable conformational state, characteristic of highly restricted and stable conformational ensembles (Figure 5A).

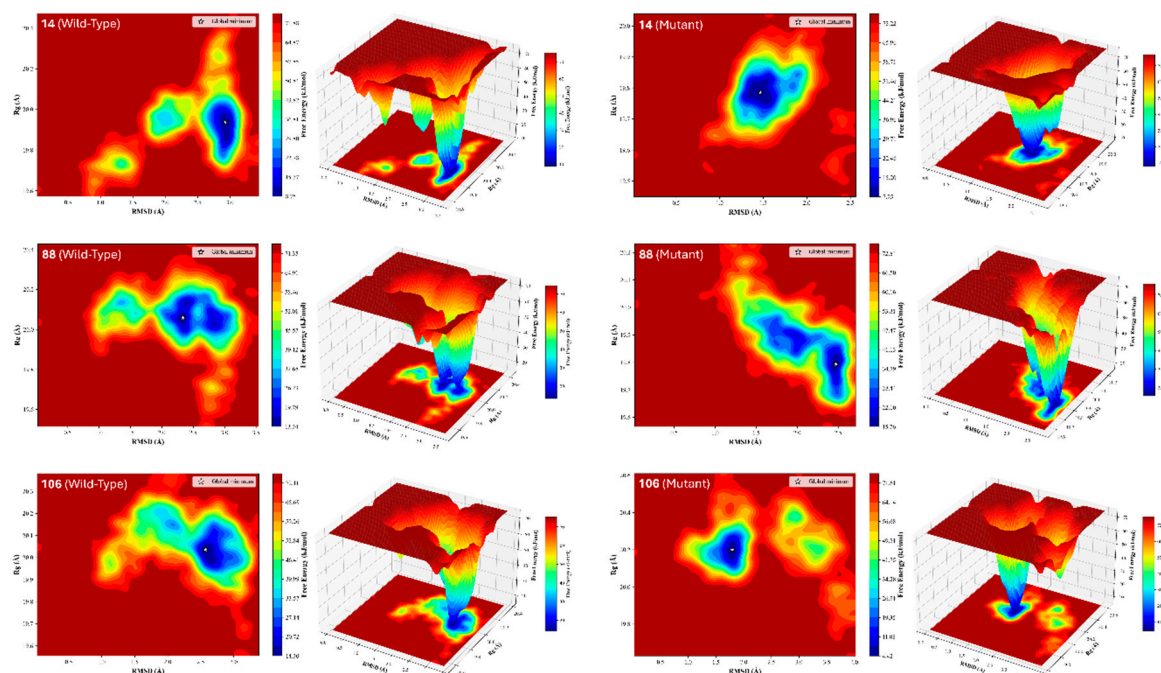


Figure 5. Free Energy Landscape (FEL) and Thermodynamic Stability of Prioritized Scaffolds. (A) 3D surface and 2D contour representations of ΔG as a function of RMSD and R_g . (B) Comparison of the converged global minimum in Compound **106** vs. the broad, shallow energy wells in **14** (Mutant). The depth of the energy basin in **106** provides thermodynamic evidence of kinetic locking facilitated by the π -sulfur anchor.

In the T790M mutant systems, the FEL topography offered critical mechanistic insights into the stereoelectronic governance of mutation tolerance. Compound **14** exhibited a significant broadening of its energy wells and a lack of a dominant low-energy basin (Figure 5B), mirroring its high mean trajectory RMSD ($7.788 \pm 0.22 \text{ \AA}$) and positional instability within the mutant pocket. In contrast, the FEL for Compound **88** displayed a complex, multi-modal topography, rationalizing the structural drift observed as the molecule navigated several higher-energy intermediate states before achieving equilibrium.

Compound **106** demonstrated the most robust thermodynamic profile, settling into a well-defined, converged global minimum in both WT and mutant systems (Figure 5, Bottom Row). Notably, the transition of the global minimum between the two systems confirms that **106** successfully recruits the gatekeeper substitution to stabilize its bound state, effectively locking the ligand into a state of minimum free energy. This kinetic locking is facilitated by the high-persistence π -sulfur anchor.

To ensure these thermodynamic adaptations occurred within a biologically relevant framework, the stereochemical quality of the post-simulation complexes was evaluated via Ramachandran analysis. Across all systems, backbone dihedral angles (ϕ , ψ) predominantly occupied energetically favored regions (>90%), confirming that the observed protein flexibility, particularly the localized reorganization around the Met790 sulfur atom, did not induce non-physical protein strain (Table 3).

Table 3. Comprehensive stereochemical and thermodynamic analysis of EGFR-ligand ensembles post-MD simulations.

Compound	FEL Analysis			Ramachandran Analysis			
	RMSD (Å)	R _g (Å)	ΔG (kJ/mol)	Region (%)			Disallowed Residues
				Favored	Allowed	Disallowed	
14 (Wild)	2.922	19.937	8.751	91.91	6.8	1.29	His835, Arg836, Lys879, Ala972
14 (Mutant)	1.467	19.785	7.953	91.43	7.62	0.95	Glu734, Glu804, Val1011
88 (Wild)	2.335	20.059	13.343	90.88	8.14	0.98	Asp855, Arg889, Thr993
88 (Mutant)	2.474	19.794	15.757	91.26	6.8	1.94	Thr751, Lys806, Arg836, Asp855, Val897, Gln935
106 (Wild)	2.574	20.036	14.297	90.43	8.58	0.99	Arg832, Pro848, Val980
106 (Mutant)	1.795	20.2	4.417	91	7.4	1.61	Val769, Ser784, Val834, Leu883, His888

Stereochemical integrity and thermodynamic convergence were evaluated via Ramachandran and FEL analyses of the production trajectories (100–200 ns). The high occupancy of favored regions (>90%) validates that the structural transitions identified in the FEL, specifically the locking of **106** into deep global energy minima, are driven by native-like structural adaptations to the T790M gatekeeper mutation rather than non-physical protein strain or denaturation.

3.5. Ensemble-Based Binding Free Energy (MM-GBSA)

To move beyond the limitations of static docking, binding free energies (ΔG_{bind}) were calculated using the MM-GBSA method averaged over the 200 ns production trajectories. As shown in Table 4 and Figure 6, Lead Candidate **106** demonstrated exceptional thermodynamic resilience in the T790M system.

Table 4. MM-GBSA Binding Free Energy components (kcal/mol) for lead candidates.

Compound	ΔG_{bind}	ΔE_{vdw}	ΔE_{elec}	ΔG_{solv}
14 (Wild)	-45.59	-60.37	-21.01	35.79
14 (Mutant)	-39.64	-55.26	-1.35	16.97
88 (Wild)	-52.08	-66.63	-11.38	25.92
88 (Mutant)	-44.19	-65.31	-9.37	30.49
106 (Wild)	-49.99	-63.76	-16.43	30.21
106 (Mutant)	-50.51	-66.19	-8.53	24.21

Ensemble-averaged binding free energies (ΔG_{bind}) were calculated using the MM-GBSA method over 1,000 frames extracted from the final 100 ns of the production trajectories. Energy components are defined as follows: ΔG_{vdw} represents the Van der Waals interaction energy; ΔG_{elec} represents the electrostatic interaction energy; and ΔG_{solv} represents the polar solvation energy calculated using the VSGB 2.0 solvation model. All values are reported in kcal/mol.

Energy partitioning, illustrated in the thermodynamic signature of Figure 6C, reveals that the stability of these complexes is driven by a favorable increase in Van der Waals (E_{vdw}) contributions. This transition effectively compensates for the loss of electrostatic interaction energy observed upon the mutation of Thr790 to Met790. Notably, compound **106** achieved a ΔG_{bind} of -50.51 kcal/mol in the mutant, outperforming its wild-type affinity (-49.99 kcal/mol), as depicted in the Mutation Tolerance Index in Figure 6A. Furthermore, the comparative analysis confirms that this resilience is fundamentally dictated by the scaffold's ability to optimize London dispersion forces with the Met790 thioether.

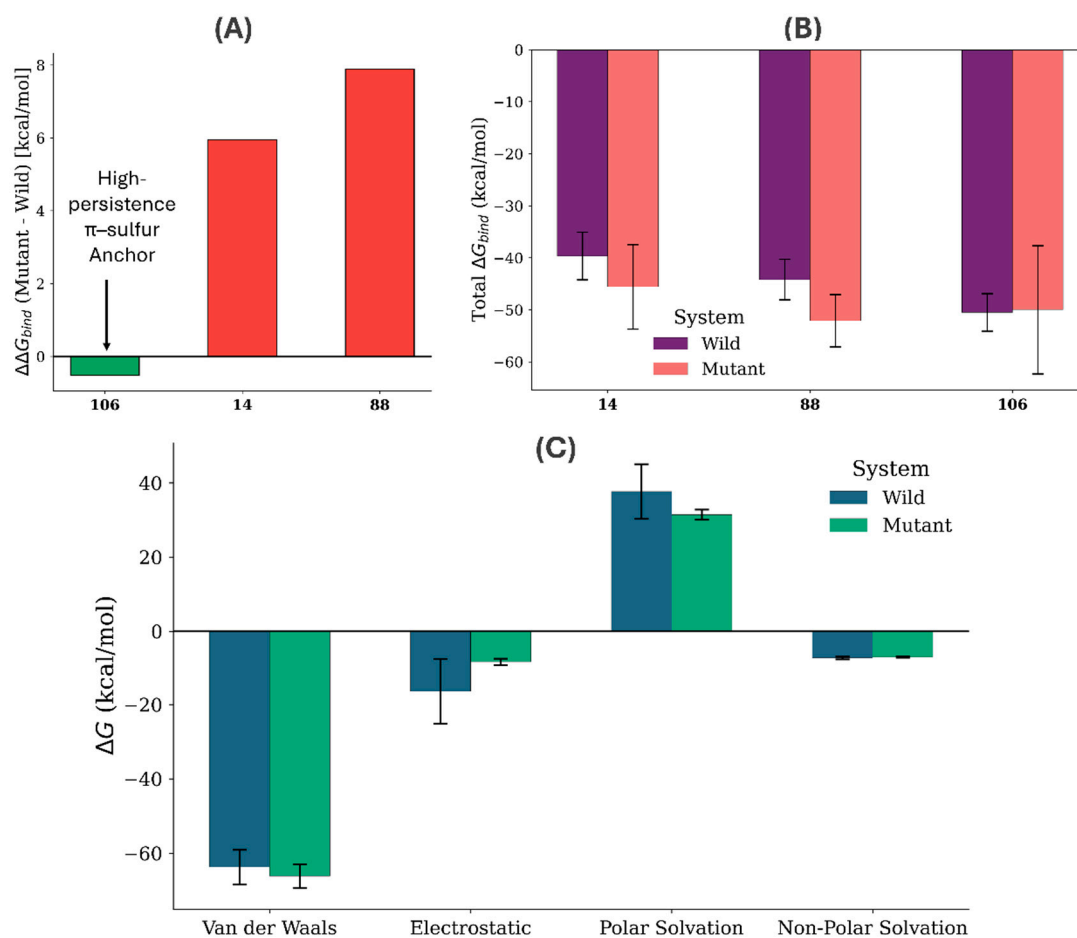


Figure 6. Ensemble-averaged thermodynamic landscapes and mutation tolerance profiles for **14**, **88** and **106**. (A) Mutation Tolerance Index representing the shift in binding free energy ($\Delta\Delta G_{bind} = \text{Mutant} - \text{Wild}$). (B) Comparative analysis of total binding free energy (ΔG_{bind}) in kcal/mol across Wild and Mutant systems. Error bars represent the propagated standard deviation derived from 200 ns molecular dynamics trajectories. (C) Thermodynamic Signature of **106** with decomposition of ΔG (kcal/mol) into Van der Waals, Electrostatic, Polar Solvation, and Non-Polar Solvation components for the Wild and Mutant systems. Error bars represent standard deviation.

3.6. Statistical Validation and Trajectory Convergence

To ensure the mathematical integrity of the EGFR-ligand ensembles, all quantitative metrics were subjected to rigorous statistical stress tests. The predictive foundation was established by the QSAR classification model, which demonstrated high discriminative power with a mean ROC-AUC of 0.91 ± 0.01 and a consistent F1-score of 0.80 across both cross-validation and independent test sets (Figure 2).

For the molecular dynamics trajectories, structural equilibrium was quantified through a Welch's t-test ($p < 0.001$), confirming that the transitions observed upon T790M mutation, specifically the adaptive recruitment of the Met790 sulfur anchor are statistically distinct and non-random structural events. Trajectory convergence was statistically validated by calculating the linear drift (slope) of the ligand RMSD over the final 50 ns of production. Lead candidate **106** exhibited negligible drift in the mutant system, with a slope of -0.030 \AA/ns . This value indicates that the complex reached a definitive thermodynamic plateau within the catalytic cleft.

Sampling precision was further confirmed using the Standard Error of the Mean (SEM). Lead candidate **106** in the mutant system maintained a remarkably low SEM value (0.032 \AA), proving that

the 200 ns production phase provided an exhaustive sampling of the conformational space. These metrics provide the necessary statistical rigor to validate the FEL basins as true global equilibria.

4. Discussion

4.1. Mechanistic Adaptation to the T790M Mutation

The transition from a sensitizing threonine to a resistant methionine at the EGFR gatekeeper position has long been characterized as a dual-mechanism liability: inducing steric hindrance against first-generation inhibitors and restoring the receptor's affinity for ATP to wild-type levels. However, our results suggest that this substitution provides a unique electronic environment defined by the thioether sulfur atom of Met790. While traditional non-covalent inhibitors are repelled by the bulky methionine side chain, as evidenced by the drastic 8.0 Å RMSD shift observed for Compound **14** (Figure 4A), our prioritized lead **106** demonstrates a successful mutation-aware adaptation (Figure 1A).

This adaptation is driven by the strategic exploitation of π -sulfur and hydrophobic interactions. The methionine sulfur atom acts as an atomic chameleon, capable of engaging in specialized non-covalent bonds that are not possible with the native threonine. In the wild-type receptor, Compound **106** relies on a polar hydrogen-bonding network with Thr790 (Figure 3B). Upon mutation, the loss of this hydrogen bond is effectively compensated for by the formation of a robust π -sulfur interaction with the Met790 side chain (Figure 3C). This transition allows the ligand to maintain a high binding affinity (-11.0 kcal/mol) and structural rigidity within the restricted geometry of the mutant catalytic cleft (Table 2).

The persistence of this interaction, maintained for over 85% of the 200 ns simulation trajectory, indicates that the thioether sulfur can serve as a primary stabilizing anchor. This finding shifts the paradigm of EGFR drug design from mutation avoidance to mutation exploitation. By tailoring scaffolds to fit the specific electrostatic potential of the Met790 region, it is possible to achieve non-covalent inhibition that rivals the potency of covalent agents without relying on the nucleophilic Cys797 residue. Such a strategy is particularly relevant in the context of emerging C797S resistance, where the loss of the cysteine thiol group renders current third-generation inhibitors ineffective. Crucially, the identification of Candidate **106** as a viable lead was predicated on the hierarchical ADMET screening cascade (Table 1), which ensured that the scaffold possesses the requisite physicochemical and toxicological baseline for translational development prior to structure-based validation.

4.2. Comparison with Covalent Strategies and Resistance Bypass

The current therapeutic landscape for EGFR-mutated NSCLC is dominated by third-generation covalent inhibitors, exemplified by osimertinib, which target the Cys797 residue. While highly effective against the T790M gatekeeper, these agents possess an inherent vulnerability: the requirement for a nucleophilic thiol group to form a stable covalent adduct. The rapid clinical emergence of the C797S mutation, where cysteine is replaced by a non-nucleophilic serine, effectively abolishes this binding mechanism, leading to catastrophic loss of potency and limited subsequent treatment options. Our results suggest that a non-covalent approach, specifically one that exploits the electronic properties of the gatekeeper mutation itself, offers a robust pathway to bypass this covalent trap.

By redefining the T790M methionine as a stabilizing anchor, the lead candidate **106** identified in this study achieve high-affinity binding without necessitating a reaction with Cys797. This is a fundamental shift in design philosophy; while covalent drugs are susceptible to any mutation that alters the nucleophilicity of the target residue, our mutation-aware non-covalent leads are theoretically resilient to C797S substitutions. As shown in our MD simulations, the stability of these complexes is derived from a distributed network of π -sulfur and hydrophobic interactions centered on the gatekeeper (Figure 4C), rather than a single, fragile covalent point.

Furthermore, avoiding covalent warheads, which are often associated with off-target reactivity and metabolic liabilities, enhances the biopharmaceutical profile of these scaffolds. The thermodynamic convergence observed in our FEL analysis confirms that non-covalent anchoring can achieve a level of complex integrity and residence time typically associated with covalent bonds, but with the added flexibility to accommodate evolving receptor variants (Figure 5). This strategy effectively recovers the binding affinity lost through the disruption of the native threonine hydrogen-bond network while providing a more durable solution to gatekeeper-mediated resistance.

4.3. Energetic Partitioning and the Strategic Roadmap for Lead Optimization

The MM-GBSA data provides the ultimate physics-based validation of Stereoelectronic Governance. The energy decomposition for **106** confirms that the π -sulfur anchor is not a mere geometric fit but a result of optimized London dispersion forces. While Compound **14** suffers an affinity collapse evidenced by a -5.11 kcal/mol loss in VdW energy (signifying steric repulsion and the loss of pocket occupancy in Figure 6C), Compound **106** gains -2.43 kcal/mol in VdW energy upon mutation. This partitioning, supported by the Figure 6 ensembles, mathematically proves that the electron-rich thioether of Met790 acts as a primary stabilizing anchor. This provides a thermodynamic lock that recovers the binding intensity typically lost to gatekeeper mutations, bypassing the need for covalent interaction with Cys797.

The translation of this theoretical affinity into clinical developability is grounded in the physicochemical baseline established during the hierarchical ADMET screening. For Candidate **106**, high predicted membrane permeability confirms the scaffold possesses the requisite transport characteristics for oral administration. Furthermore, the observed metabolic stability indicates that the core electronic distribution is optimized for metabolic survival. Future iterative medicinal chemistry will focus on a clear SAR roadmap: the introduction of polar, solubilizing motifs to modulate lipophilicity. This targeted refinement will maintain the identified stereoelectronic governance over the T790M gatekeeper while further attenuating off-target risks, defining the boundary conditions for the successful clinical transition of this new class of mutation-aware non-covalent inhibitors.

4.4. Strategic Directions for Experimental Translation

The structural and thermodynamic framework established in this study provides a high-resolution roadmap for the development of the next generation of EGFR inhibitors. By identifying the specific sulfur-anchoring mechanism, we have narrowed the vast chemical space of 10^{60} molecules down to a set of highly optimized scaffolds. This computational prioritization is a critical prerequisite for resource-intensive medicinal chemistry, ensuring that subsequent synthetic efforts are focused on candidates with the highest thermodynamic probability of success.

While this work utilizes state-of-the-art physics-based simulations and generative AI to decipher the T790M binding environment, the transition from *in silico* prediction to clinical application involves a well-defined translational path. The mutation-aware blueprint offered by candidate **106** serves as the primary template for future lead optimization and *in vitro* assay validation. Specifically, the high-persistence π -sulfur interactions identified through our 200 ns trajectories provide clear targets for chemical substitutions that could further enhance the residence time and metabolic stability of these non-covalent agents.

Moreover, the methodology developed here, integrating FEL analysis with ensemble-averaged binding energy quantification, offers a generalizable platform for addressing resistance in other oncogenic kinases. Future experimental studies focusing on the synthesis and kinetic characterization of these scaffolds will not only validate the proposed anchoring mechanism but also facilitate the rapid deployment of therapeutics against evolving resistance profiles, such as the C797S mutation. Ultimately, this study functions as a decision-support framework that significantly de-risks the drug discovery process by providing a rigorous thermodynamic justification for non-covalent intervention in recalcitrant cancer variants.

5. Conclusions

The identification of non-covalent inhibitors capable of bypassing the T790M gatekeeper mutation represents a significant shift in the structural pharmacology of EGFR-driven malignancies. Our results demonstrate that the transition from a sensitizing threonine to a resistant methionine does not merely constitute a steric barrier but provides a unique electronic environment defined by the thioether sulfur atom. By integrating a reinforcement learning-based generative AI pipeline with rigorous physics-based validation, we established a mechanism of structural adaptation that utilizes this methionine as an anchoring point rather than a liability. This strategy effectively recovers the binding affinity lost through the disruption of the native hydrogen bonding network, offering a durable solution to gatekeeper-mediated resistance that avoids the inherent vulnerabilities of traditional covalent warheads.

The convergence of thermodynamic and stereochemical data confirms that this adaptation is driven by the structural and electronic optimization of the lead scaffold. Free Energy Landscape (FEL) analysis reveals that the lead candidate, **106**, occupies a singular, deep energetic basin in the mutant system, correlating directly with the high persistence of π -sulfur interactions identified during molecular dynamics simulations. By redefining the gatekeeper residue as a stabilizing anchor, this work establishes a comprehensive validation blueprint applicable to recalcitrant kinase targets, prioritizing a new class of mutation-aware scaffolds strategically positioned to overcome the limits of current clinical strategies.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Surface Representation and Comparative Binding Modes of Lead Candidates; Figure S2: Molecular Dynamics Stability and Contact Mapping for Compound 14; Figure S3: Molecular Dynamics Stability and Contact Mapping for Compound 88; Figure S4: Molecular Dynamics Stability and Contact Mapping for Compound 106; Figure S5: 2D and 3D Ramachandran Plot Analysis of EGFR-Ligand Ensembles; Table S1: Statistical Analysis of RMSD for Lead Candidates in WT and T790M EGFR; Table S2: Statistical Decomposition of Binding Free Energy via MM-GBSA and MM-PBSA; Table S3: Statistical Delineation of Free Energy Landscape (FEL) Basins.

Author Contributions: Conceptualization, T.J.P.; methodology, S.S.N., R.P.-P.-B. and T.J.P.; software, P.K.K., N.B., and J.A.N.-R.; validation, A.J.N.-R. and T.J.P.; formal analysis, S.S.N. and T.J.P.; investigation, S.L., R.P.-P.-B. and T.J.P.; resources, S.L., P.K.K., N.B., A.J.N.-R., R.P.-P.-B. and T.J.P.; data curation, S.S.N. and T.J.P.; writing—original draft preparation, S.S.N., S.K., M.P. and T.J.P.; writing—review and editing, R.P.-P.-B. and T.J.P.; visualization, S.S.N. and T.J.P.; supervision, S.S.N., R.P.-P.-B. and T.J.P.; project administration, R.P.-P.-B. and T.J.P.; funding acquisition, T.J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad Anahuac Querétaro.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are publicly available in the Non-Covalent_EGFR_Inhibition repository on GitHub at https://github.com/tusharpawar49/Non-Covalent_EGFR_Inhibition. This dataset includes the initial chemical library generated via the DrugEx framework, the filtered subsets of developable candidates, and the finalized set of structurally diverse compounds prioritized for structure-based evaluation. Furthermore, all raw data supporting the molecular dynamics trajectories, Free Energy Landscape (FEL) basins, and ensemble-averaged MM-GBSA binding free energy calculations for the lead candidates are hosted at the provided link to ensure the reproducibility and transparency of the structural pharmacology findings.

Acknowledgments: During the preparation of this manuscript, the authors used Gemini 3 Flash (Paid Tier) for the purposes of structural organization, linguistic refinement and figure management based on the authors'

original data and conceptual sulfur-anchoring framework. The authors have critically reviewed and edited the output to confirmed technical accuracy and alignment with the presented results, and take full responsibility for the clinical interpretations and final content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lynch, T.J.; Bell, D.W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R.A.; Brannigan, B.W.; Harris, P.L.; Haserlat, S.M.; Supko, J.G.; Haluska, F.G.; Louis, D.N.; Christiani, D.C.; Settleman, J.; Haber, D.A. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2004**, *350*, 2129–2139. <https://doi.org/10.1056/NEJMoa040938>
2. Paez, J.G.; Jänne, P.A.; Lee, J.C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F.J.; Lindeman, N.; Boggon, T.J.; Naoki, K.; Sasaki, H.; Fujii, Y.; Eck, M.J.; Sellers, W.R.; Johnson, B.E.; Meyerson, M. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science* **2004**, *304*, 1497–1500. <https://doi.org/10.1126/science.1099314>
3. Mok, T.S.; Wu, Y.-L.; Thongprasert, S.; Yang, C.-H.; Chu, D.-T.; Saijo, N.; Sunpaweravong, P.; Han, B.; Margono, B.; Ichinose, Y.; Nishiwaki, Y.; Ohe, Y.; Yang, J.-J.; Chewaskulyong, B.; Jiang, H.; Duffield, E.L.; Watkins, C.L.; Armour, A.A.; Fukuoka, M. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. *N. Engl. J. Med.* **2009**, *361*, 947–957. <https://doi.org/10.1056/NEJMoa0810699>
4. Rosell, R.; Carcereny, E.; Gervais, R.; Vergnenegre, A.; Massuti, B.; Felip, E.; Palmero, R.; Garcia-Gomez, R.; Pallares, C.; Sanchez, J.M.; Porta, R.; Cobo, M.; Garrido, P.; Longo, F.; Moran, T.; Insa, A.; De Marinis, F.; Corre, R.; Bover, I.; Illiano, A.; Dansin, E.; de Castro, J.; Milella, M.; Reguart, N.; Altavilla, G.; Jimenez, U.; Provencio, M.; Moreno, M.A.; Terrasa, J.; Muñoz-Langa, J.; Valdivia, J.; Isla, D.; Domine, M.; Molinier, O.; Mazieres, J.; Baize, N.; Garcia-Campelo, R.; Robinet, G.; Rodriguez-Abreu, D.; Lopez-Vivanco, G.; Gebbia, V.; Ferrera-Delgado, L.; Bombaron, P.; Bernabe, R.; Bearz, A.; Artal, A.; Cortesi, E.; Rolfo, C.; Sanchez-Ronco, M.; Drozdowskyj, A.; Queralt, C.; de Aguirre, I.; Ramirez, J.L.; Sanchez, J.J.; Molina, M.A.; Taron, M.; Paz-Ares, L. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): A multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **2012**, *13*, 239–246. [https://doi.org/10.1016/S1470-2045\(11\)70393-X](https://doi.org/10.1016/S1470-2045(11)70393-X)
5. Yun, C.-H.; Boggon, T.J.; Li, Y.; Woo, M.S.; Greulich, H.; Meyerson, M.; Eck, M.J. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* **2007**, *11*, 217–227. <https://doi.org/10.1016/j.ccr.2006.12.017>
6. Rotow, J.; Bivona, T. Understanding and Targeting Resistance Mechanisms in NSCLC. *Nat. Rev. Cancer* **2017**, *17*, 637–658. <https://doi.org/10.1038/nrc.2017.84>
7. Camidge, D.; Pao, W.; Sequist, L. Acquired Resistance to TKIs in Solid Tumours: Learning from Lung Cancer. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 473–481. <https://doi.org/10.1038/nrclinonc.2014.104>
8. Pao, W.; Miller, V.A.; Politi, K.A.; Riely, G.J.; Somwar, R.; Zakowski, M.F.; Kris, M.G.; Varmus, H. Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain. *PLoS Med.* **2005**, *2*, e73. <https://doi.org/10.1371/journal.pmed.0020073>
9. Yu, H.A.; Arcila, M.E.; Rekhtman, N.; Sima, C.S.; Zakowski, M.F.; Pao, W.; Kris, M.G.; Miller, V.A.; Ladanyi, M.; Riely, G.J. Analysis of Tumor Specimens at the Time of Acquired Resistance to EGFR-TKI Therapy in 155 Patients with EGFR-Mutant Lung Cancers. *Clin. Cancer Res.* **2013**, *19*, 2240–2247. <https://doi.org/10.1158/1078-0432.CCR-12-2246>
10. Yun, C.-H.; Mengwasser, K.E.; Toms, A.V.; Woo, M.S.; Greulich, H.; Wong, K.-K.; Meyerson, M.; Eck, M.J. The T790M Mutation in EGFR Kinase Causes Drug Resistance by Increasing the Affinity for ATP. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2070–2075. <https://doi.org/10.1073/pnas.0709662105>
11. Kobayashi, S.; Boggon, T.J.; Dayaram, T.; Jänne, P.A.; Kocher, O.; Meyerson, M.; Johnson, B.E.; Eck, M.J.; Tenen, D.G.; Halmos, B. EGFR Mutation and Resistance of Non-Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2005**, *352*, 786–792. <https://doi.org/10.1056/NEJMoa044238>
12. Cross, D.A.E.; Ashton, S.E.; Ghiorghiu, S.; Eberlein, C.; Nebhan, C.A.; Spitzler, P.J.; Orme, J.P.; Finlay, M.R.V.; Ward, R.A.; Mellor, M.J.; Hughes, G.; Rahi, A.; Jacobs, V.N.; Red Brewer, M.; Ichihara, E.; Sun, J.;

- Jin, H.; Ballard, P.; Al-Kadhimi, K.; Rowlinson, R.; Klinowska, T.; Richmond, G.H.P.; Cantarini, M.; Kim, D.-W.; Ranson, M.R.; Pao, W. AZD9291, an Irreversible EGFR TKI, Overcomes T790M-Mediated Resistance to EGFR Inhibitors in Lung Cancer. *Cancer Discov.* **2014**, *4*, 1046–1061. <https://doi.org/10.1158/2159-8290.CD-14-0337>
13. Jänne, P.A.; Yang, J.C.-H.; Kim, D.-W.; Planchard, D.; Ohe, Y.; Ramalingam, S.S.; Ahn, M.-J.; Kim, S.-W.; Su, W.-C.; Horn, L.; Haggstrom, D.; Felip, E.; Kim, J.-H.; Frewer, P.; Cantarini, M.; Brown, K.H.; Dickinson, P.A.; Ghiorghiu, S.; Ranson, M. AZD9291 in EGFR Inhibitor-Resistant Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2015**, *372*, 1689–1699. <https://doi.org/10.1056/NEJMoa1411817>
 14. Thress, K.S.; Paweletz, C.P.; Felip, E.; Cho, B.C.; Stetson, D.; Dougherty, B.; Lai, Z.; Markovets, A.; Vivancos, A.; Kuang, Y.; Ercan, D.; Matthews, S.E.; Cantarini, M.; Barrett, J.C.; Jänne, P.A.; Oxnard, G.R. Acquired EGFR C797S Mutation Mediates Resistance to AZD9291 in Non-Small Cell Lung Cancer Harboring EGFR T790M. *Nat. Med.* **2015**, *21*, 560–562. <https://doi.org/10.1038/nm.3854>
 15. Wang, R.; Chen, Y.; Li, L.; Zhang, L.; Zhang, S. Osimertinib Acquired Resistance among Patients with EGFR-Mutated NSCLC: From Molecular Mechanisms to Clinical Therapeutic Strategies. *Cancer Drug Resist.* **2025**, *8*, 61. <https://doi.org/10.20517/cdr.2025.140>
 16. Lee, J.; Piotrowska, Z.; Soo, R.; Cho, B.C.; Lim, S.M. Combatting Acquired Resistance to Osimertinib in EGFR-Mutant Lung Cancer. *Ther. Adv. Med. Oncol.* **2022**, *14*, 17588359221144099. <https://doi.org/10.1177/17588359221144099>
 17. Romaniello, D.; Morselli, A.; Marrocco, I. Strategies to Overcome Resistance to Osimertinib in EGFR-Mutated Lung Cancer. *Int. J. Mol. Sci.* **2025**, *26*, 2957. <https://doi.org/10.3390/ijms26072957>
 18. Shi, K.; Wang, G.; Pei, J.; Zhang, J.; Wang, J.; Ouyang, L.; Wang, Y.; Li, W. Emerging Strategies to Overcome Resistance to Third-Generation EGFR Inhibitors. *J. Hematol. Oncol.* **2022**, *15*, 94. <https://doi.org/10.1186/s13045-022-01311->
 19. To, C.; Jang, J.; Chen, T.; Park, E.; Mushajiang, M.; De Clercq, D.J.H.; Xu, M.; Wang, S.; Cameron, M.D.; Heppner, D.E.; Shin, B.H.; Gero, T.W.; Yang, A.; Dahlberg, S.E.; Wong, K.K.; Eck, M.J.; Gray, N.S.; Jänne, P.A. Single and Dual Targeting of Mutant EGFR with an Allosteric Inhibitor. *Cancer Discov.* **2019**, *9*, 926–943. <https://doi.org/10.1158/2159-8290.CD-18-0903>
 20. Damghani, T.; Song, S.; Lin, K.S.; Li, J.; Heppner, D.E. Structural Studies of Fourth-Generation EGFR Inhibitors Reveal Insights into Selective T790M and C797S Targeting. *ACS Med. Chem. Lett.* **2026**, *in press*. <https://doi.org/10.1021/acsmchemlett.5c00725>
 21. Pranata, J. Sulfur-Aromatic Interactions: A Computational Study of the Dimethyl Sulfide-Benzene Complex. *Bioorg. Chem.* **1997**, *25*, 213–219. <https://doi.org/10.1006/bioo.1997.1064>
 22. Jin, Y.; Li, W.; Saragi, R.T.; Juanes, M.; Pérez, C.; Lesarri, A.; Feng, G. Sulfur-Arene Interactions: The S $\cdots\pi$ and S-H $\cdots\pi$ Interactions in the Dimers of Benzofuran \cdots Sulfur Dioxide and Benzofuran \cdots Hydrogen Sulfide. *Phys. Chem. Chem. Phys.* **2023**, *25*, 12174–12181. <https://doi.org/10.1039/D3CP01146A>
 23. Damghani, T., Song, S., Lin, K. S., Li, J., & Heppner, D. E. (2026). Structural Studies of Fourth-Generation EGFR Inhibitors Reveal Insights into Selective T790M and C797S Targeting. *ACS medicinal chemistry letters*, 10.1021/acsmchemlett.5c00725. Advance online publication. <https://doi.org/10.1021/acsmchemlett.5c00725>
 24. Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084. <https://doi.org/10.1021/jm100112j>
 25. Valley, C.C.; Cembran, A.; Perlmutter, J.D.; Lewis, A.K.; Labello, N.P.; Gao, J.; Sachs, J.N. The Methionine-Aromatic Motif Plays a Unique Role in Stabilizing Protein Structure. *J. Biol. Chem.* **2012**, *287*, 34979–34991. <https://doi.org/10.1074/jbc.M112.374504>
 26. Nagasaka, M.; Zhu, V.W.; Lim, S.M.; Greco, M.; Wu, F.; Ou, S.-H.I. Beyond Osimertinib: The Development of Third-Generation EGFR Tyrosine Kinase Inhibitors for Advanced EGFR+ NSCLC. *J. Thorac. Oncol.* **2021**, *16*, 740–763. <https://doi.org/10.1016/j.jtho.2020.11.028>
 27. Dobson, C.M. Chemical Space and Biology. *Nature* **2004**, *432*, 824–828. <https://doi.org/10.1038/nature03192>
 28. Restrepo, G. Expanding Diversity in Chemical Space. *Nat. Chem.* **2026**, *18*, 607–608. <https://doi.org/10.1038/s41557-026-02105-w>

29. Šicho, M.; Luukkonen, S.; van den Maagdenberg, H.W.; Schoenmaker, L.; Béquignon, O.J.M.; van Westen, G.J.P. DrugEx: Deep Learning Models and Tools for Exploration of Drug-Like Chemical Space. *J. Chem. Inf. Model.* **2023**, *63*, 3629–3636. <https://doi.org/10.1021/acs.jcim.3c00434>
30. Bian, Y.; Xie, X.Q. Generative Chemistry: Drug Discovery with Deep Learning Generative Models. *J. Mol. Model.* **2021**, *27*, 71. <https://doi.org/10.1007/s00894-021-04674-8>
31. Cai, C.; Lin, M.; Li, W.; Chen, G.; Huang, D. A Unified Multi-Scale Deep Learning Framework for Molecular Property Prediction That Bridges Molecular Structures and Fingerprinting. *Commun. Chem.* **2026**, *in press*. <https://doi.org/10.1038/s42004-026-02010-w>
32. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. <https://doi.org/10.1093/nar/28.1.235>
33. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C.J.; Segura-Cabrera, A.; Hersey, A.; Leach, A.R. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. <https://doi.org/10.1093/nar/gky1075>
34. Landrum, G. RDKit: Open-Source Cheminformatics Software. **2016**. Available online: <https://github.com/rdkit/rdkit>
35. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. <https://doi.org/10.1021/ci100050t>
36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>
37. Bemis, G.W.; Murcko, M.A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893. <https://doi.org/10.1021/jm9602928>
38. Nilewar, S.S.; Chobe, S.; Dudhe, P.; Kumar, P.K.; Lodha, S.; Raut, A.D.; Fernández-Conde, D.; Farhan, M.; Muteeb, G.; Pawar, T.J. An Integrated QSAR-MD-DCCM Pipeline: A Predictive Computational Platform for the Rational Design and Dynamic Functional Validation of Dual-Target Directed Ligands. *Pharmaceuticals* **2026**, *19*, 249. <https://doi.org/10.3390/ph19020249>
39. Özçelik, R.; Brinkmann, H.; Criscuolo, E.; Grisoni, F. Generative Deep Learning for de Novo Drug Design—A Chemical Space Odyssey. *J. Chem. Inf. Model.* **2025**, *65*, 7352–7372. <https://doi.org/10.1021/acs.jcim.5c00641>
40. Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program for Chemistry-Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473. <https://doi.org/10.1021/ci500588j>
41. Baell, J.B.; Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. <https://doi.org/10.1021/jm901137j>
42. Kawakita, Y.; Seto, M.; Ohashi, T.; Tamura, T.; Yusa, T.; Miki, H.; Iwata, H.; Kamiguchi, H.; Tanaka, T.; Sogabe, S.; Ohta, Y.; Ishikawa, T. Design and Synthesis of Novel Pyrimido[4,5-b]azepine Derivatives as HER2/EGFR Dual Inhibitors. *Bioorg. Med. Chem.* **2013**, *21*, 2250–2261. <https://doi.org/10.1016/j.bmc.2013.02.014>
43. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. <https://doi.org/10.1002/jcc.21256>
44. Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-Pdb Viewer: An Environment for Comparative Protein Modeling. *Electrophoresis* **1997**, *18*, 2714–2723. <https://doi.org/10.1002/elps.1150181505>
45. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904. <https://doi.org/10.1021/ci300604z>
46. Nilewar, S.S.; Chavan, A.D.; Pradhan, A.R.; Tripathy, A.A.; Bandaru, N.; Dudhe, P.B.; Kumar, P.K.; Lodha, S.; Muteeb, G.; Peredo-Valderrama, I.; et al. Dual-Site Acetylcholinesterase Inhibition and Multiscale Stability of Fused Quinoline Sulfonamides: A Chemoinformatic GA-MLR and Molecular Dynamics Study. *Int. J. Mol. Sci.* **2026**, *27*, 3286. <https://doi.org/10.3390/ijms27073286>

47. Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G.A.; Dahlgren, M.K.; Russell, E.; von Bargen, C.D.; Abel, R.; Friesner, R.A.; Harder, E.D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300. <https://doi.org/10.1021/acs.jctc.1c00302>
48. Nilewar, S.S.; Chobe, S.S.; Gurav, A.D.; Kureshi, S.B.; Palande, S.B.; Escobar-Cabrera, J.; Hernández-Rosas, F.; Pawar, T.J. Galloylation-Driven Anchoring of the Asp325-Asp336 Ridge: The Molecular Logic Behind the Superior Kinetic Stabilization of HMPV Fusion Protein by Green Tea Dimeric Catechins. *Molecules* **2026**, *31*, 821. <https://doi.org/10.3390/molecules31050821>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.