Evaluating the Docking Performance of Erlotinib within the EGFR Tyrosine Kinase Domain

Joudy Allam, Haya Mazyad

Department of Computer Science & Mathematics, Lebanese American University

BIF513: Computational Biology

Dr. Joseph Rebehmed

December 2, 2024

Docking Performance: EGFR and Erlotinib

I. Biological Significance Of EGFR-AQ4 Complex

The epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase essential for regulating cellular processes such as proliferation, differentiation, and survival. When activated by its ligands, EGFR undergoes autophosphorylation, triggering downstream signaling pathways. In cancers such as non-small-cell lung cancer (NSCLC), pancreatic cancer, and glioblastomas, EGFR is often overexpressed or mutated, leading to continuous activation of these pathways. This results in uncontrolled cell growth, inhibition of apoptosis, and enhanced angiogenesis and metastasis, making EGFR a key therapeutic target.

Erlotinib (Figure 1.a), marketed as Tarceva, is a small-molecule inhibitor belonging to the 4-anilinoquinazoline (AQ4) class. It binds reversibly to the ATP-binding pocket of the EGFR tyrosine kinase domain, preventing autophosphorylation and halting the downstream signaling necessary for tumor progression. This targeted mechanism specifically disrupts mitogenic and anti-apoptotic pathways in cancer cells, offering a more precise therapeutic approach compared to traditional chemotherapy. Erlotinib is primarily used to treat metastatic NSCLC with specific EGFR mutations and pancreatic cancer.

The 4-anilinoquinazoline scaffold, the structural foundation of erlotinib, confers high specificity and affinity for EGFR by allowing it to compete effectively with ATP at the kinase domain. This scaffold also supports the development of derivative inhibitors to overcome resistance mutations, such as EGFR T790M, enhancing its therapeutic potential. Erlotinib's small-molecule nature provides several advantages over monoclonal antibodies, including deeper tumor penetration, the ability to target cytoplasmic receptor domains, and significantly lower production costs, making it a versatile and accessible option in cancer treatment.

Docking Performance: EGFR and Erlotinib

The crystal structure of the erlotinib-EGFR complex (Figure 1.b), resolved at 2.6 Å, highlights the precise interactions that stabilize the inactive conformation of EGFR. Erlotinib occupies the ATP-binding pocket with key hydrogen bonds and hydrophobic interactions, providing insights into its inhibitory mechanism. This complex not only underscores the therapeutic efficacy of EGFR inhibitors but also offers a structural framework for designing improved drugs. Understanding the erlotinib-EGFR interaction exemplifies the promise of precision medicine in cancer therapy.

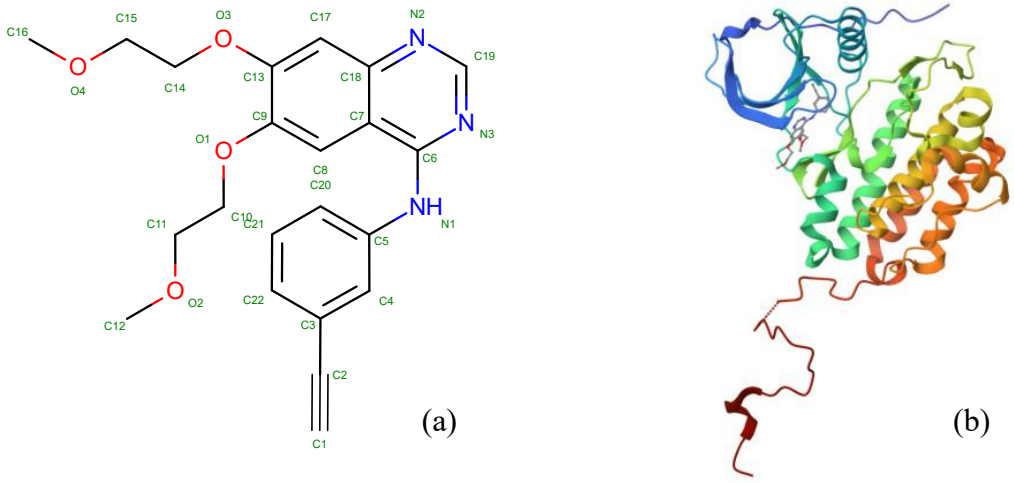


Figure 1: (a) 2D structure of 4-anilinoquinazoline (AQ4), also known as [6,7-bis(2-methoxy-ethoxy)quinazoline-4-yl]-(3-ethynylphenyl)amine, the chemical scaffold of erlotinib. (b) Crystal structure of the epidermal growth factor receptor (EGFR) tyrosine kinase domain in complex with the 4-anilinoquinazoline inhibitor erlotinib (PDB ID: 1M17).

Artificial Separation of Ligand and Macromolecule

To separate the ligand from the macromolecule, a Linux-based approach was utilized to process the PDB file. The working directory was adjusted to the location of the input file (`1M17.pdb`), and the following command was executed:

```
awk '/^HETATM/ && $4 == "AQ4" {print > "ligand.pdb"} /^ATOM/ {print > "macromolecule.pdb"}' 1M17.pdb
```

- `/^HETATM/ && $4 == "AQ4"`: Filters lines beginning with `HETATM` where the residue name in column 4 matches AQ4, identifying the ligand. This ensures that only the ligand, AQ4, is extracted, excluding other HETATM entries such as water molecules or other non-ligand heteroatoms. The filtered lines are saved in a file named `ligand.pdb`.
- `/^ATOM/`: Filters lines starting with `ATOM`, corresponding to the macromolecule (receptor). These lines are saved in a file named `macromolecule.pdb`.
- The input file, `1M17.pdb`, is the source containing both the macromolecule and ligand information.

The output files are the following:

- `ligand.pdb`: Contains the structural data for the ligand, AQ4.
- `macromolecule.pdb`: Contains the structural data for the macromolecule.

This approach ensures a clean separation of the ligand (AQ4) and receptor (EGFR) components from the original PDB file, facilitating their independent preparation for subsequent docking simulations. The ligand file will be used for structural analysis and docking, while the receptor file will be prepared for grid placement and docking evaluations.

II. Preparation of Receptor and Ligand Files for Docking

Using the regex command `^ATOM.{12}B.*` in Sublime Text, alternate locations for 14 atoms in the receptor (EGFR) were identified and subsequently removed. The complex's PDB files indicate the presence of missing atoms and chain breaks, as illustrated in Figure 2.a. However, these atoms are located far from the ligand binding site, rendering their absence negligible for the docking process, therefore no corrections were necessary.

Crystallographic water molecules were automatically removed during receptor preparation using the Linux command previously applied for file separation.

Next, hydrogen atoms were added to the receptor using AutoDock Tools, and Kollman and Gasteiger charges were assigned. For the ligand (AQ4), the root was identified (Figure 2.b), and its rotatable bonds were determined, totaling 10 torsions (Figure 2.c) with 14 aromatic carbons found.

Docking Performance: EGFR and Erlotinib

Finally, both the receptor and the ligand files were converted to `.pdbqt` format.

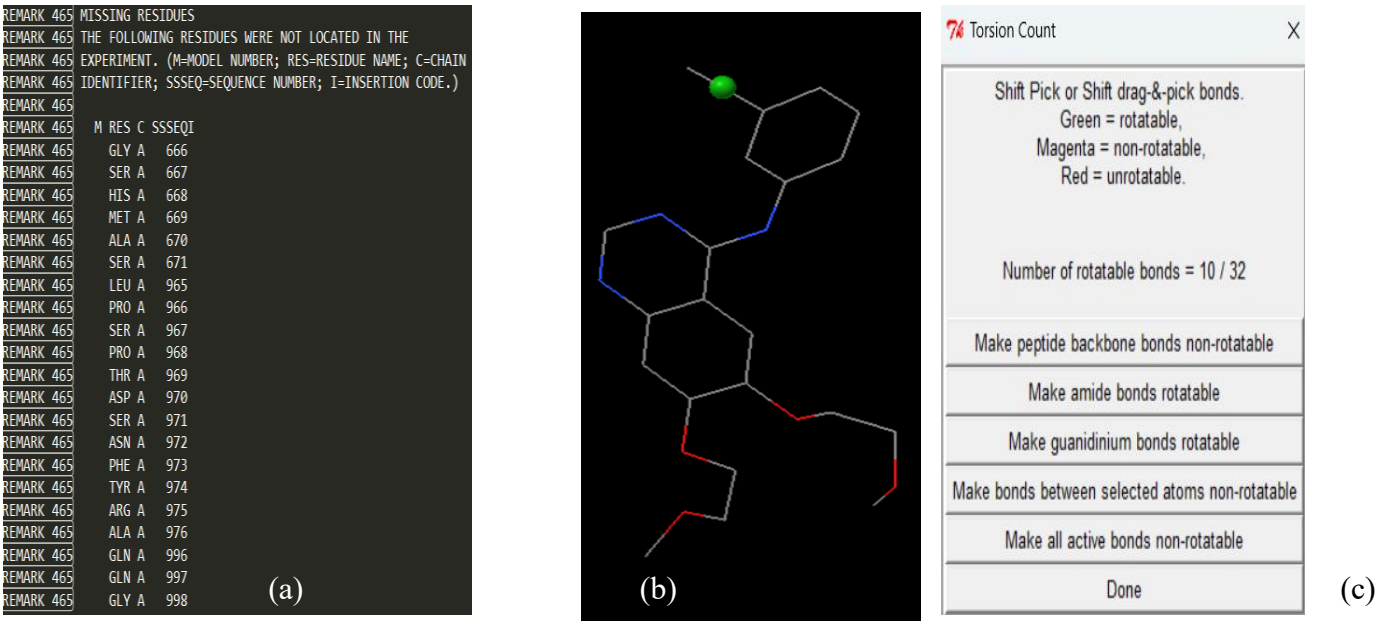


Figure 2: (a) Missing residues in the EGFR receptor as listed in REMARK 465 of the 1M17 PDB file. (b) Molecular structure of the ligand (AQ4) with the root atom identified and represented by the green sphere. The structure is color-coded: red represents oxygen atoms, blue represents nitrogen atoms, and gray represents carbon atoms, illustrating the composition and connectivity of the molecule. (c) Rotatable bonds of the ligand (AQ4) identified using AutoDockTools with 10 out of 32 bonds classified as rotatable.

III. Defining the Binding Site and Computing Interaction Energy Maps

To compute the interaction energy maps for the AQ4-EGFR complex, a grid box was defined to cover the receptor's binding site. This ensures that all possible binding interactions between the AQ4 and EGFR are explored while maintaining computational efficiency. The grid box dimensions (x, y, z) are (60, 68, 58) and the center of the grid box is (23.543, 9.849, 59.407). A grid spacing of 0.375 Å was used, balancing precision in the interaction calculations with computational efficiency. These dimensions were chosen to fully enclose the binding site, allowing the ligand to interact with all relevant residues.

Figure 3 shows the grid box placement around the receptor's binding site. The blue box highlights the area covered by the grid, illustrating how it effectively surrounds the receptor's

active site. The dimensions and spacing allow for accurate mapping of the ligand's interactions with the receptor.

By defining these parameters, the grid box ensures that the docking simulations are precise and encompass the entire binding site while minimizing computational overhead.

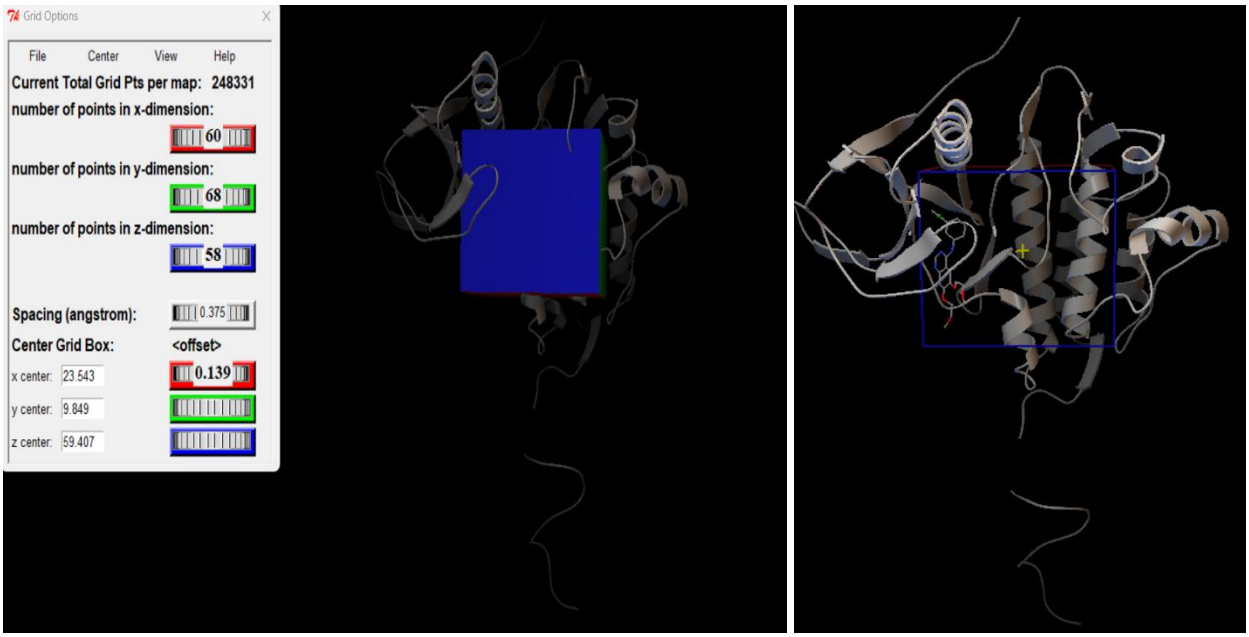


Figure 3: Grid placement around the receptor's binding site, illustrating the defined dimensions, center, and spacing used for docking simulations.

IV. Docking Simulation Using the Lamarckian Genetic Algorithm

The docking of the ligand within the receptor's binding site was performed using the Lamarckian Genetic Algorithm (LGA). LGA effectively explores the ligand's conformational space, combining a global search through a genetic algorithm with local search refinements for accurate pose prediction. The docking process was guided by the pre-calculated interaction energy maps, generated during the grid preparation step. These maps represented the energetic contributions of various atom types in the ligand (A NA C OA) across the receptor's binding site, serving as the basis for scoring the predicted poses.

Docking Performance: EGFR and Erlotinib

For this docking study, the search parameters were carefully configured to balance computational efficiency with accuracy. Specifically, a population size of 150, 100 generations, and 2500000 (medium) energy evaluations were used.

The docking process involved the use of 2 key files: the docking parameter file (DPF) used to perform the docking and the docking log file (DLG) downloaded at the end.

- `aq4.dpf`: essential for defining the parameters and settings required for the docking simulation. It contains detailed information, including the receptor and ligand filenames, the pre-calculated grid maps, the search algorithm (Lamarckian Genetic Algorithm), and its parameters, and the grid box information, which defined the region of the receptor's binding site explored during the docking process.
- `aq4.dlg`: provides a comprehensive record of the docking process, including the energies associated with each conformation, the ranked list of docking results, with detailed information on the ligand's position, orientation, and interactions within the receptor's binding pocket for each pose.

V. Comparative Analysis of Docked and Experimental Poses

i. Characterization of Ligand Atom Types and Charge

The 4 atom types in the ligand are **A** (Aromatic Carbon), **NA** (Acceptor Nitrogen), C (Aliphatic Carbon), and OA (Acceptor Oxygen). These atom types are crucial for defining the chemical and electronic properties of the ligand, which influence its interactions with the binding site. The total charge of the ligand is **-0.001e**, which is effectively neutral.

ii. Torsional Degrees of Freedom and Binding-Associated Free Energy Changes

AQ4 has exactly 10 torsional degrees of freedom, all active. These torsions are between the following pairs of atoms:

1. C5_2517 and N1_2518
2. N1_2518 and C6_2519
3. C9_2522 and O1_2523
4. O1_2523 and C10_2524
5. C10_2524 and C11_2525
6. C11_2525 and O2_2526
7. C13_2528 and O3_2529
8. O3_2529 and C14_2530
9. C14_2530 and C15_2531
10. C15_2531 and O4_2532

The estimated loss of torsional free energy upon binding is **+2.983 kcal/mol**. This value represents the energetic penalty associated with the reduction in conformational flexibility of the ligand as it transitions from a free state to a bound state within the receptor's binding pocket.

iii. Determination of the Best Estimated Free Energy of Binding

The best estimated Free Energy of Binding between the ligand and the receptor is **-7.23 kcal/mol**. This value was obtained from **LGA Run 67,** which produced the pose with the highest predicted binding affinity.

This energy represents the sum of all interaction energies between the ligand and receptor, including van der Waals forces, hydrogen bonding, electrostatics, and desolvation effects (-10.21), offset by any torsional penalties (+2.98). A more negative

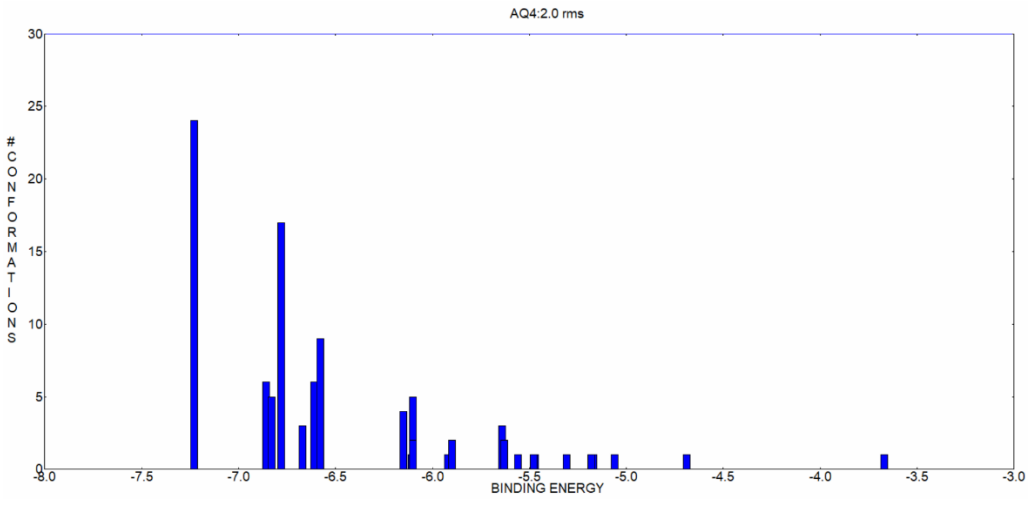
binding energy indicates stronger binding affinity, suggesting that this pose is likely the most favorable conformation for the ligand-receptor complex.

iv. Clustering Analysis of Docked Ligand Conformations

The ligand poses were clustered based on structural similarity and energy, as determined by their Root Mean Square Deviation (RMSD) and binding energy values. Poses with RMSD values ≤ 2 Å relative to each other were grouped within the same cluster, while poses with RMSD values > 2 Å were classified into different clusters. The most stable pose, defined as the one with the lowest binding energy, was used as the seed for each cluster. This clustering ensures that poses within each cluster are conformationally and energetically similar, while those across clusters are both structurally and energetically distinct.

A total of 25 clusters were obtained, with each cluster representing a distinct group of ligand poses. The frequency of poses in each cluster, out of 100 generated runs, is provided in the column **"Num in Clus"** in the accompanying table. The clustering results are further visualized in the histogram shown below, which displays the distribution of conformations across different binding energy ranges. Each bar

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus
1	-7.23	67	-6.43	24
2	-6.86	80	-6.05	6
3	-6.83	92	-5.91	5
4	-6.78	11	-6.17	17
5	-6.67	65	-5.79	3
6	-6.61	78	-5.84	6
7	-6.58	96	-6.05	9
8	-6.15	56	-5.72	4
9	-6.11	61	-6.11	1
10	-6.10	63	-5.67	5
11	-6.10	99	-5.78	2
12	-5.92	81	-5.92	1
13	-5.90	48	-5.52	2
14	-5.64	19	-5.09	3
15	-5.63	12	-5.31	2
16	-5.56	42	-5.56	1
17	-5.48	83	-5.48	1
18	-5.47	25	-5.47	1
19	-5.31	71	-5.31	1
20	-5.18	30	-5.18	1
21	-5.17	88	-5.17	1
22	-5.17	37	-5.17	1
23	-5.06	77	-5.06	1
24	-4.69	73	-4.69	1
25	-3.67	87	-3.67	1



represents the number of ligand poses within a specific binding energy interval, grouped into clusters based on RMSD ≤ 2 Å.

v. Evaluation of Ligand Fit and Binding Pocket Interactions

Docked Ligand Poses in the Receptor's Binding Pocket

The docked ligand poses of AQ4 fit exceptionally well within the EGFR receptor's binding pocket. The top-ranked conformation, selected based on its lowest binding energy and clustering analysis, demonstrates high complementarity with the receptor. Its fit ensures minimal steric clashes while optimizing interactions critical for stability.

Key metrics, such as allowed overlaps (0.6) and hydrogen bond overlap reduction (0.4), confirm that the docking process was driven by the complementarity of the ligand's shape and properties with the receptor's binding site. Furthermore, the absence of high-energy clashes strengthens the validity of the docking results.

Interactions Between the Top-Ranked Conformation and the Receptor:

The top-ranked conformation of the ligand shows a good fit, with a low energy score of -7.23 which indicates no clashes at any point in the complex. This suggests an optimal interaction between the ligand and the receptor, where the docking is driven by complementarity in shape and properties. Specifically, the interaction is characterized by various stabilizing forces, including hydrogen bonds, hydrophobic interactions, electrostatic interactions, and the absence of π-π stacking and π-cation interactions. To better visualize and analyze these interactions, Chimera was used to illustrate and display the detailed binding between the ligand and the receptor.

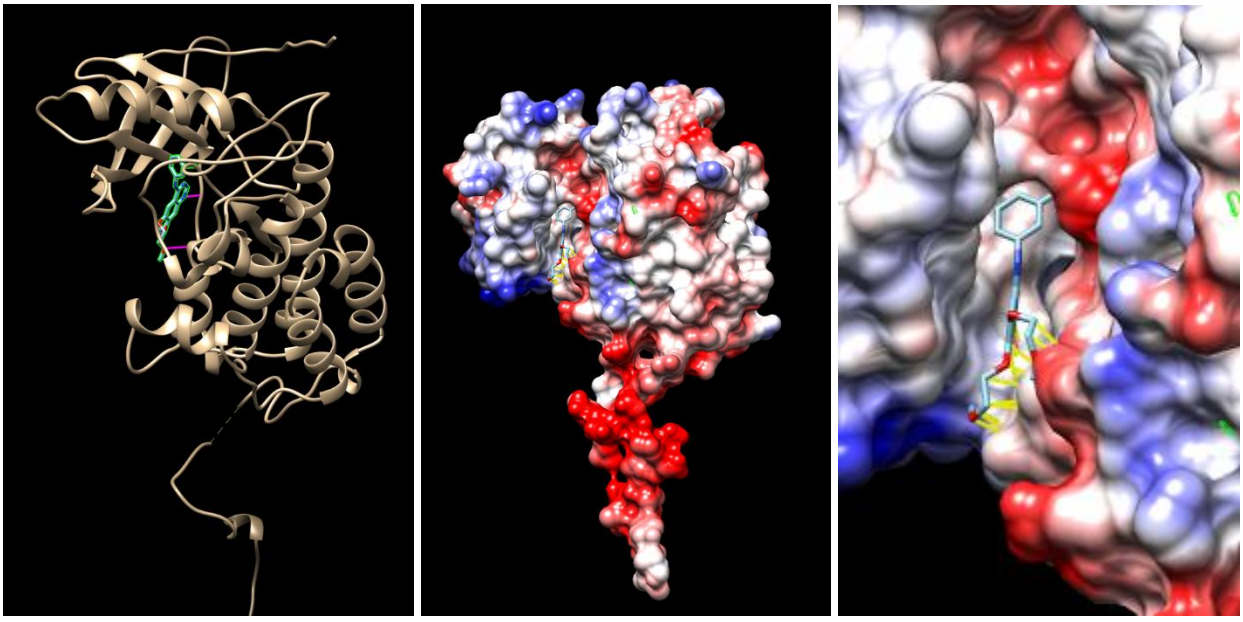
Docking Performance: EGFR and Erlotinib

- **Shape:** The ligand's top-ranked conformation fits well into the receptor's binding pocket, with no steric clashes or hindrance, also clear in Figure 4c. The optimal orientation of the ligand within the pocket ensures that all key interactions are maximized, and the shape complementarity between the ligand and receptor stabilizes the overall complex. This perfect fit contributes to the favorable low energy of the system.
- **Hydrogen Bonds:** Hydrogen bonds, represented in pink in Figure 4a, play a crucial role in stabilizing the ligand-receptor complex. Specifically, AQ4's nitrogen atom (N2) forms a hydrogen bond with the backbone nitrogen of MET769 at a distance of 2.876 Å, and its oxygen atom (O2) forms another hydrogen bond with the backbone nitrogen of CYS773 at a distance of 2.711 Å. These hydrogen bonds are key in ensuring the ligand maintains its optimal positioning within the receptor's binding pocket, contributing significantly to the specificity and stability of the interaction.
- **Hydrophobic Interactions:** Hydrophobic interactions, indicated in yellow in Figure 4c, involve close contact between the ligand's non-polar regions and hydrophobic residues of the receptor. These interactions are essential for stabilizing the ligand in the binding pocket. Specifically, AQ4's C19 interacts with ALA719 at a distance of 3.063 Å, while AQ4's C1 aligns with MET742's sulfur atom at a distance of 3.194 Å. These interactions contribute to the ligand-receptor complex by ensuring a snug, energetically favorable fit. The non-polar residues of both the receptor and the ligand interact through van der Waals forces, minimizing steric hindrance and enhancing the stability of the binding.
- **Electrostatic Interactions:** Electrostatic interactions further stabilize the complex by facilitating the binding between oppositely charged regions of the ligand and receptor. AQ4's carbonyl and amine groups interact with the charged residues ASP831 and

GLU738 of the receptor. These Coulombic interactions help anchor the ligand in place, providing additional stability and specificity to the complex. The electrostatic complementarity between the ligand and receptor enhances the binding affinity, making the interaction stronger and more stable.

- **Absence of π-π and π-Cation Interactions:** The analysis shows no π-π stacking or π-cation interactions between the ligand and receptor. Despite the absence of these interactions, the binding is still strong and stable due to the dominance of other interaction types, such as hydrogen bonds, hydrophobic contacts, and electrostatic interactions. These interactions are sufficient to maintain the ligand's optimal fit within the receptor's binding pocket, ensuring a stable and energetically favorable complex.

The interaction between the top-ranked conformation of AQ4 and the receptor is primarily driven by hydrogen bonds, hydrophobic interactions, and electrostatic forces, with a total of **70 contacts** observed.



Docking Performance: EGFR and Erlotinib

Figure 4:*(a) Hydrogen bonds formed between AQ4 and the receptor binding site, with pink lines representing the bonds. Key residues involved include MET769 and CYS773.*
(b) Surface representation of the receptor showing electrostatic complementarity in the binding pocket, highlighting areas of positive (blue) and negative (red) charge.
(c) Close-up view of the binding pocket illustrating key hydrophobic interactions (yellow) and electrostatic interactions stabilizing the ligand-receptor complex.

vi. Comparative Analysis of Docked and Experimental Ligand Structures

The **Reference RMSD** is a measure of the structural similarity between a predicted (docked) and experimental (reference) structure, indicating the average distance between corresponding atoms. Here, the docked poses show strong similarity to the experimental structure, especially the top-ranked poses. The best pose (Rank 1, Sub-Rank 1), with the most favorable energy (-7.23 kcal/mol), has a Reference RMSD of 1.95 Å, which is within the commonly accepted threshold of 2 Å for good agreement with the experimental conformation. Other high-ranking poses, such as Sub-Rank 2 (1.86 Å) and Sub-Rank 10 (1.75 Å), also exhibit close alignment with the experimental structure.

The superposition of the best docking-generated conformation (blue) and the experimental conformation (pink) of AQ4 in the EGFR binding pocket (Figure 5a) reveals notable differences in orientation and positioning. While both conformations reside within the same binding site, the docked pose exhibits a slight shift and rotation compared to the experimental pose, indicating variations in how the ligand interacts with the protein. These differences may result in altered interaction profiles, such as changes in hydrogen bonding (in the experimental ligand only 1 hydrogen bond was shown in comparison to the 2 in the best docking pose), hydrophobic interactions, or steric clashes with nearby residues.

Docking Performance: EGFR and Erlotinib

However, lower-ranked poses display higher RMSD values, such as Rank 2, Sub-Rank 1 (3.37 Å) and Rank 3, Sub-Rank 1 (3.91 Å), indicating poorer alignment. These deviations may be due to alternative binding modes or inaccuracies in predicting interactions within a flexible binding site.

One important observation is run 76, which corresponds to Rank 1, Sub-Rank 4 (Figure 5b). It exhibits the lowest Reference RMSD (1.39 Å) and thus aligns closest to the experimental structure among all docked poses. Despite this high level of structural similarity, it is not ranked as the top pose. This can be attributed to the fact that the ranking system prioritizes a combination of factors, including binding energy and clustering RMSD, rather than relying solely on Reference RMSD. While Run 76's structural alignment is excellent, its Binding Energy (-6.88 kcal/mol) is slightly less favorable than the top-ranked pose, which has a binding energy of -7.23 kcal/mol.

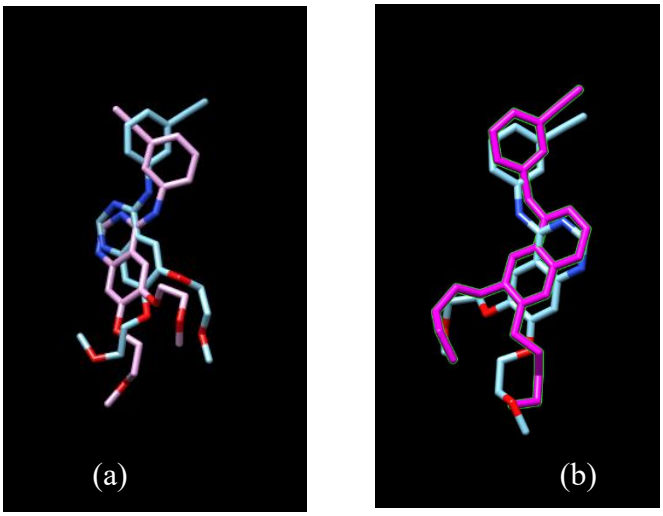


Figure 5:*(a) Superposition of the best-docked conformation (blue) and the experimental conformation (pink) of AQ4 in the EGFR binding site, highlighting differences in orientation and positioning.*
(b) Superposition of Run 76 docking conformation (pink) and experimental conformation (blue) of AQ4 in the EGFR binding site, illustrating variations in ligand alignment and interactions

Additional Alternates

Another docking simulation was conducted using the B alternates of the ligand instead of the A alternates employed in the initial run. While 14 alternates were found, only 8 of them are located within the binding site, and these are the conformations that will influence the docking results. The B alternates represent an alternative set of atomic conformations within the ligand structure, enabling a comparative analysis of binding poses, energies, and interactions. This approach aimed to determine whether these conformations produced more favorable binding energies or distinct binding modes.

When the macromolecule was superimposed with the A and B alternates, some noticeable changes in the ligand's binding orientation and positioning within the receptor’s active site were observed (Figure 6a). These changes, highlighted in the accompanying image, demonstrate how the ligand's conformational flexibility impacts its interaction with the receptor. The comparison of the A and B alternates reveals variations in binding pose that could potentially affect the ligand's binding affinity and specificity.

The new docking run, conducted with the B alternates of the ligand, presents several differences compared to the previous docking run using the A alternates. The top-ranked pose in this docking (Rank 1) has a binding energy of -7.27 kcal/mol, slightly better than the previous best binding energy of -7.23 kcal/mol. However, the average binding energy across all poses is also higher, with the mean binding energy for this docking at -6.27 kcal/mol, compared to -6.67 kcal/mol from the previous run.

Docking Performance: EGFR and Erlotinib

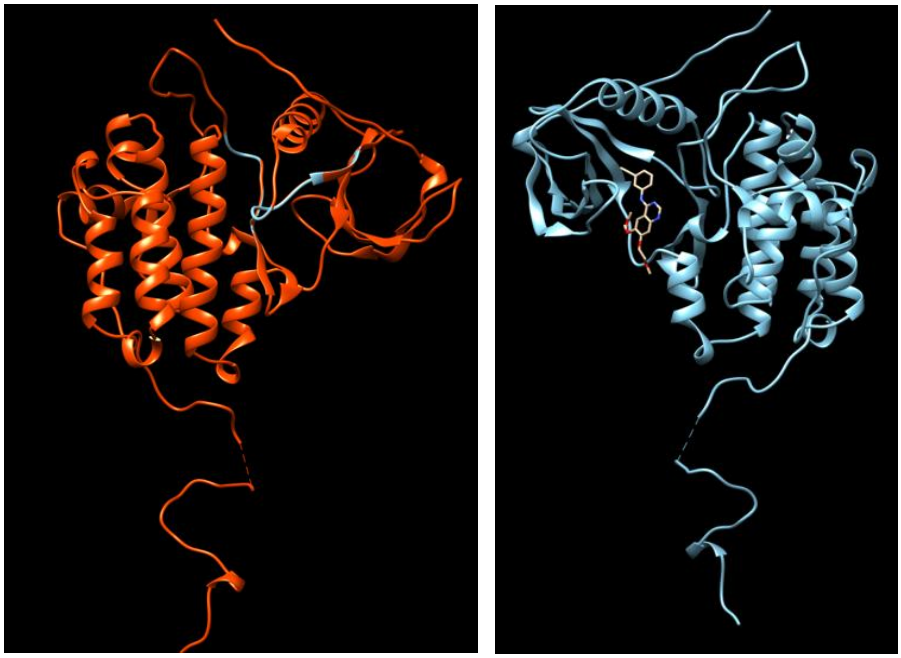


Figure 6:*(a) Superimposition of the macromolecule with A and B alternates, showing changes in binding pose.*

(b) Binding of the best-docked pose with macromolecule with B alternates.

References

1 Bank, R. P. D. (n.d.). 1M17: Epidermal growth factor receptor tyrosine kinase domain with 4-anilinoquinazoline inhibitor Erlotinib. RCSB PDB. https://www.rcsb.org/structure/1M17

Bank, R. P. D. (n.d.-b). Aq4 ligand summary page. RCSB PDB. https://www.rcsb.org/ligand/AQ4

Stamos, J., Sliwkowski, M. X., & Eigenbrot, C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. The Journal of biological chemistry, 277(48), 46265–46272. (n.d.). https://doi.org/10.1074/jbc.M207135200

Welcome to AutoDock. AutoDock. https://autodock.scripps.edu/

UCSF Chimera Home Page. (n.d.). https://www.rbvi.ucsf.edu/chimera/