Investigating Potential EGFR Kinase Inhibitors for Lung Cancer Treatment by Evaluating Protein-Ligand Docking

Casey Nguyen, SID 06420690, caseyhn@stanford.edu

Introduction

The epidermal growth factor receptor (EGFR) is involved in cell growth and survival. EGFR also plays an essential role in cancer growth and proliferation as mutations to the EGFR gene can lead to the evasion of apoptosis. Mutated forms of the EGFR genes and protein are associated with certain types of cancer, such as non-small cell lung cancer. EGFR tyrosine kinase inhibitors are highly effective against lung cancers with mutations to the EGFR gene and protein. I aim to assess the docking of osimertinib (also known by its trade name Tagrisso), an oral EGFR Kinase inhibitor, with the EGFR Kinase protein domain. Additionally, by evaluating the docking for other drug options, I seek to gain insight on alternative treatments that would provide flexibility if cancer builds acquired resistance to a currently administered drug.

Background

EGFR encodes the EGFR protein, a receptor tyrosine kinase part of the ErbB family. The EGFR protein is transmembrane, containing both an extracellular and intracellular domain. It binds to ligands to receive and transmit signals to the cell. When the receptor binds to a ligand, it dimerizes or attaches to another EGFR receptor, activating a receptor complex. The complex triggers signaling pathways that promote cell growth, division, and survival.

Mutations in the EGFR gene and protein activate the receptor signal complex, enabling the cell to grow and divide uncontrollably, leading to cancer. EGFR mutations, such as EGFRm and EGFR T790M (a secondary point mutation substituting methionine with threonine at amino

acid position 790), have been identified in non-small cell lung cancers where the mutations cause uncontrolled cell growth in the lung tissue. EGFR kinase inhibitors, such as osimertinib, treat cancer by competing with ATP for binding to EGFR kinase, suppressing the activation of the signal complex and, thus, cell growth and division. While lung cancer cells are initially highly-sensitive to osimertinib, they develop resistance and become less sensitive to the drug over time. Because of acquired drug resistance, it is crucial to have multiple drug options available such as gefitinib and erlotinib, which are also EGFR kinase inhibitors.

Additionally, because the five-year survival rate for lung cancer (18.6%) is lower than other types of cancers such as colorectal (64.5%), breast (89.6%), and prostate (98.2%), it is vital to have a variety of options for drug treatments. If an individual's lung cancer builds acquired resistance to a drug, it would be helpful to know what other drugs could still be effective against the disease.

Methods

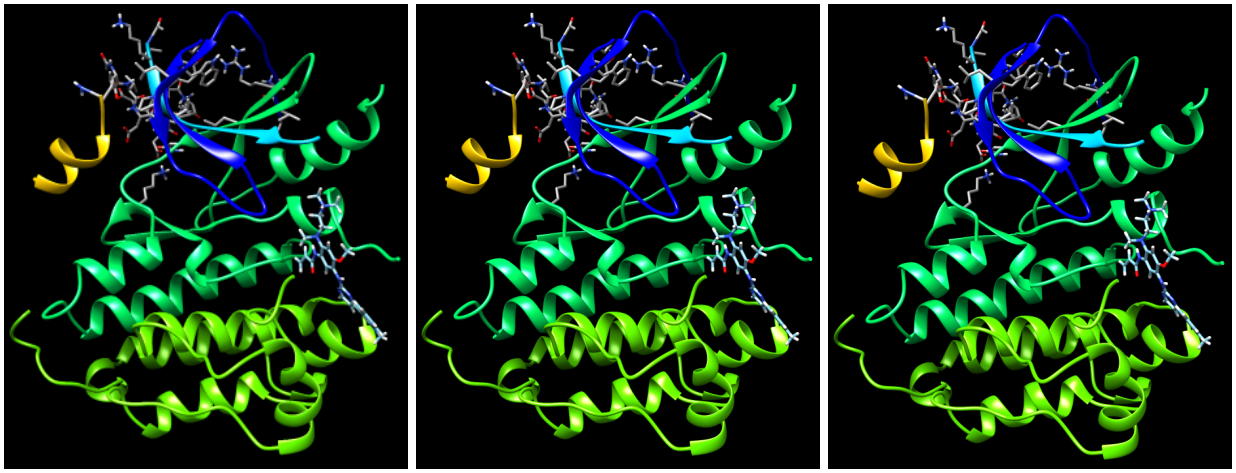
To evaluate the docking of osimertinib, SwissDock was used to predict protein-ligand docking. The EGFR Kinase domain was docked with osimertinib, a drug approved for EGFR Kinase inhibition. The pdb file for EGFR Kinase was retrieved for docking using SwissDock's search functionality for target selection. SwissSimilarity was used to find drugs with similar chemical structures to osimertinib. The chemical structure of osimertinib was retrieved using SwissDock's search functionality for ligand selection. The predicted binding modes were returned after docking completion, and UCSF Chimera was launched to visualize the modes. A preset (Interactive 1 (ribbons)) was applied to the image to distinguish the chains and protein structures. The three poses with the best FullFitness values were recorded and the pose with the lowest

estimated ΔG value. These steps for protein-ligand docking were repeated for gefitinib and erlotinib.

Results and Analysis

The following table shows the protein-ligand docking results between EGFR Kinase and osimertinib for the three poses with the best FullFitness values.

Pose	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
1	-1690.40	-9.20
2	-1689.89	-9.16
3	-1689.89	-9.16



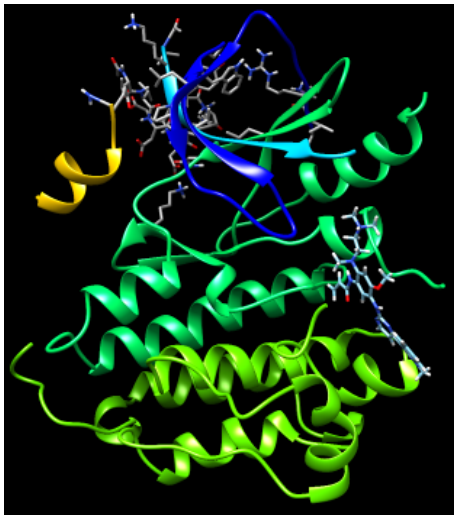
Pose 1 Pose 2 Pose 3

The resulting images for all three poses are nearly identical, with minor differences in the docking angle of the ligand. The similar structures explain the close FullFitness and estimated ΔG values between the three poses. The most favorable binding pose for EGFR Kinase docking with osimertinib had a FullFitness of -1690.40 kcal/mol and an estimated ΔG of -9.20 kcal/mol. It was surprising to me that the pose with the lowest estimated ΔG value (-10.19 kcal/mol) had a lower FullFitness value (-1667.64 kcal/mol). Though, the FullFitness values are still very close.

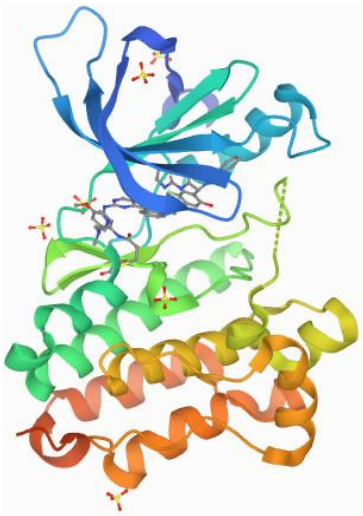
The negative Gibbs' free energy value for the three poses indicates that the docking of osimertinib to EGFR Kinase is favorable for those poses. This favorable interaction suggests that osimertinib may be able to compete with ATP in binding to EGFR Kinase, inhibiting cell growth and division.

To discover any complications in implementing the protein-ligand docking, I wanted to compare the predicted docking structure to the crystal structure of the EGFR Kinase-osimertinib complex. Any significant differences between the predicted and crystal structures would inform me of issues in the preparation or docking process.

EGFR Kinase-osimertinib complex



Predicted docking structure of complex



Crystal structure of complex

While the structure of EGFR Kinase in the predicted docking structure and crystal structure of the EGFR Kinase-osimertinib complex are similar, the ligand docking sites are noticeably different. As part of SwissDock's search feature for finding PDB native structures was nonfunctional, it was not possible to select specific chains from a structure from the Protein Database. Because of this, any ligands attached to the EGFR Kinase in the PDB source file were

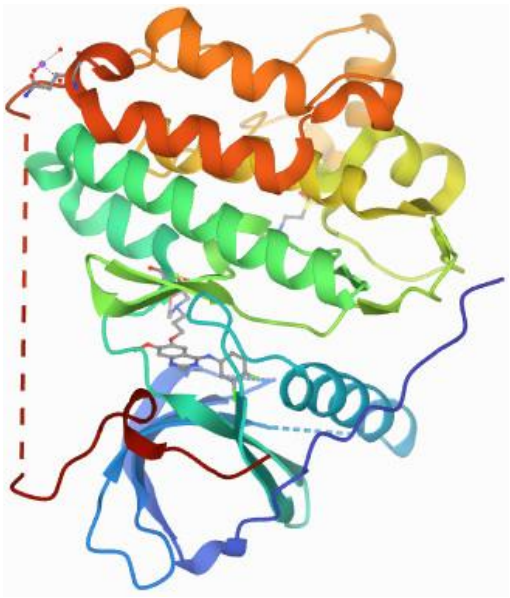
likely still attached, possibly affecting docking interactions. Until SwissDock fixes its search feature for protein chains from a PDB code, it would be challenging to prepare the protein for docking within the web server thoroughly.

The following table shows the protein-ligand docking results between EGFR Kinase and gefitinib for the three poses with the best FullFitness values.

Pose	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
1	-1680.46	-9.09
2	-1678.68	-8.96
3	-1677.56	-8.75

Additionally, the pose of the EGFR Kinase-gefitinib complex with the lowest Gibbs' free energy value had a FullFitness value of -1677.33 kcal/mol and an estimated ΔG value of -9.67 kcal/mol.

EGFR Kinase-gefitinib complex



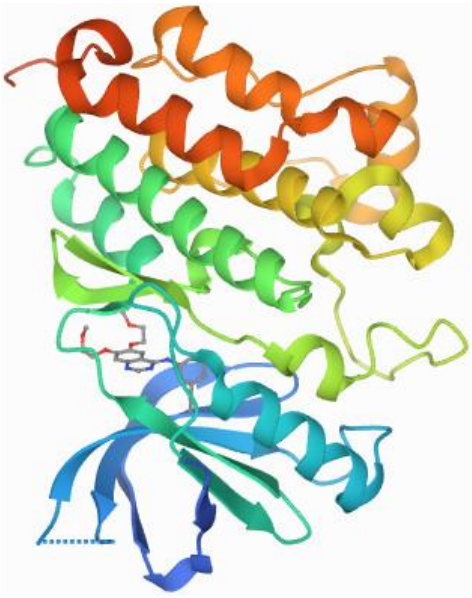
Predicted docking structure of complex Crystal structure of complex

The following table shows the protein-ligand docking results between EGFR Kinase and erlotinib for the three poses with the best FullFitness values.

Pose	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
1	-1676.87	-8.71
2	-1676.87	-8.71
3	-1676.83	-8.70

Additionally, the pose of the EGFR Kinase-gefitinib complex with the lowest Gibbs' free energy value had a FullFitness value of -1676.36 kcal/mol and an estimated ΔG value of -9.55 kcal/mol.

EGFR Kinase-erlotinib complex



Predicted docking structure of complex Crystal structure of complex

The FullFitness values of the docking structure for gefitinib and erlotinib with EGFR Kinase are similar to the values for osimertinib, and the estimated ΔG are also similar. I also

learned that the pose with the lowest Gibbs' free energy does not always have the best FullFitness value from analyzing my results. However, the ΔG value with the best FullFitness value is typically similar to the ΔG value with minimized Gibbs' free energy. As the Gibbs' free energy values are all negative for the top poses of each drug, it suggests that the drugs have favorable interactions with the receptor protein. Osimertinib, gefitinib, and erlotinib have different structures yet all have a favorable binding affinity with EGFR Kinase. It may be possible to substitute one drug for another if cancer cells develop acquired resistance to one of the drugs.

Further Work

Because SwissDock has not yet fixed its search feature for protein chains from a PDB code, I would like to prepare the protein structure for docking through a software program such as the Molecular Operating Environment (MOE). Through MOE, I would select protein chains of interest from a PDB file and further process it for docking through energy minimization, protonation (adding nonpolar hydrogen atoms), and fixing bond angles and lengths.

Alternative protein-ligand docking tools such as AutoDock Vina, Rosetta Dock, and GLIDE may also provide additional features not available in SwissDock. For example, using Rosetta Dock, I could more easily prepare the target protein with commands and implement functions I create myself using Python.

I would also experiment with PyMol's mutagenesis feature to introduce mutations to the protein of interest and observe changes in docking behavior with ligands. Because cancer cells can develop acquired resistance to drugs, it would be insightful to observe what EGFR mutations decrease the receptor's binding affinity with specific drugs. Monitoring changes to FullFitness

values for drugs with various protein mutations could inform alternative drug options when cancer develops resistance to one drug.

Additionally, I seek to discover other potential cancer drug candidates by searching for molecules that share similar chemical structures with osimertinib and evaluating their physicochemical properties. One way to approach this challenge is to use SwissSimilarity, SwissTargetPrediction, and SwissADME in addition to SwissDock. I could find similar molecules to approved drugs, predict what molecules or structures they would target in the human body, and assess pharmacokinetic properties, druglike nature and medicinal chemistry friendliness. Alternatively, instead of using SwissSimilarity to search for other potential drug candidates, I could scan through research papers on investigational drugs and assess their docking behavior and properties using computational tools.

References

AstraZeneca. (n.d.). Tagrisso Nurse. Retrieved December 3, 2021, from https://www.tagrissonurse.com/patient-identification.html?source=tag_n_h_538&umedium=cpc&uadpub=google&ucampaign=2021tagrissohcpbranded_nursegeneral&ucreative=branded_moa_ex&uplace=osimertinibmoa&outcome=hcp&cmpid=1&gclid=Cj0KCQiA-qGNBhD3ARIsAO_o7ykiXYH0oUK-BAumScYqnw9Y1z2DP6mOPVbHB7vfva1-Zh5hEwL-A2kaApjlEALw_wcB&gclsrc=aw.ds#moa.

Lung cancer fact sheet. American Lung Association. (2020, May 27). Retrieved December 3, 2021, from https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/resource-library/lung-cancer-fact-sheet#:~:text=The%20lung%20cancer%20five%2Dyear,and%20prostate%20(98.2%20percent).&text=The%20five%2Dyear%20survival%20rate,localized%20(within%20the%20lungs).

Mitsudomi, T., & Yatabe, Y. (2009). Epidermal growth factor receptor in relation to tumor development: EGFR Gene and cancer. *FEBS Journal*, *277*(2), 301–308. https://doi.org/10.1111/j.1742-4658.2009.07448.x

Suda, K., Onozato, R., Yatabe, Y., & Mitsudomi, T. (2009). EGFR T790M mutation: A double role in lung cancer cell survival? *Journal of Thoracic Oncology*, *4*(1), 1–4. https://doi.org/10.1097/jto.0b013e3181913c9f

U.S. National Library of Medicine. (2020, August 18). *EGFR gene: Medlineplus genetics*. MedlinePlus. Retrieved December 3, 2021, from https://medlineplus.gov/genetics/gene/egfr/.