Full Length Research Article
Advancements in Life Sciences – International Quarterly Journal of Biological Sciences

ARTICLE INFO

Date Received: 17/09/2024;
Date Revised: 13/12/2024;
Available Online: 31/12/2024;

**Author's Affiliation:**
Faculty of Nursing, King Abdulaziz University, Jeddah – Saudi Arabia

***Corresponding Author:**
Fatmah Hazza Alsharif
Email: Falsharif@kau.edu.sa

**How to Cite:**
Alsharif FH (2025). Identification of *Allium cepa* compounds as Promising Inhibitors Against Lung Cancer: An in-Silico Study. Adv. Life Sci. 12(1): 163-167.

**Keywords:**
Lung cancer, EGFR, Allium cepa, bioactive compounds, virtual screening

Open Access

# Identification of *Allium cepa* compounds as Promising Inhibitors against Lung Cancer: An in-Silico Study

Fatmah Hazza Alsharif*

## Abstract

**Background:** Lung cancer is one of the primary causes of cancer-related deaths, and treatment options for advanced-stage disease remain restricted. Overexpression of the epidermal growth factor receptor (EGFR) has been linked to the development of certain cancers. Double-mutated EGFR is an important oncogenic protein in many lung cancer instances. *Allium cepa*, a common condiment herb, is known for its medical and pharmacological benefits.

**Methods:** The bioactive compound of *A. cepa* was obtained from the LOTUS database in 'sdf' format, and then converted into 'pdbqt' format. The prepared compounds library was screened against the double-mutated EGFR using the insilico tool PyRx 0.8 to determine the binding conformations with the lowest binding energies.

**Result:** Eighteen compounds were found to strongly bind with the EGFR protein and have lower binding energy than the cocrystal ligand, with the top five hits being LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270. The Asn842, Asp855, Lys745, Met790, Gln791, Leu792, Met793, Ala743, Leu844, Leu718, Val726, Thr854, and Phe723 residues of EGFR were important in binding to these hit compounds. In addition, these compounds have good drug-like properties.

**Conclusion:** The compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 can be used as EGFR inhibitors to manage lung cancer. However, additional experimental studies are required to validate these compounds as EGFR inhibitors.

## Introduction

Lung cancer is the most common cause of cancer-related death in North America and other affluent countries. The 2020 special report on lung cancer describes it as the most often diagnosed cancer and the leading cause of cancer mortality in Canada, accounting for more deaths than colorectal, pancreatic, and breast cancers combined. In 2020, an estimated 30,000 Canadians will be diagnosed with lung cancer, with 21,000 expected to die from it. Globally, the cancer burden is anticipated to double by 2050, with lung cancer remaining the most common [1].

Epidermal growth factor receptor (EGFR) is a transmembrane receptor tyrosine kinase protein found in normal epithelial, mesenchymal, and neurogenic cells. Its overexpression is associated with the development of a variety of malignancies, including non-small cell lung cancer (NSCLC) [2,3]. EGFR expression in NSCLC is linked with shorter survival, increased lymph node metastases, and reduced chemosensitivity [4,5]. Targeting activating mutations in the EGFR tyrosine kinase has resulted in great success in the treatment of lung cancer, a common malignancy and the main cause of cancer-related fatalities [6]. This study focuses on the double-mutated EGFR protein, which is an important therapeutic target in lung cancer [7-9].

Modern therapeutic options are based on a disease's biological origin. Historically, natural remedies were employed to treat or reduce illnesses. Advances in research and technology led to the discovery of macromolecules involved in illnesses, guiding chemists to design and synthesis more effective bioactive chemicals. However, bringing a drug to market requires multiple steps, hurdles, and significant resources, with annual expenses reaching $2.8 billion between 2009 and 2018 [10]. To solve these economic and time problems, new techniques are required. Computer-Aided Drug Design (CADD) has become a crucial tool for drug discovery and development. Both academics and pharmaceutical firms utilize CADD to find and optimize bioactive molecules. CADD has been used to find or optimize many medications at various stages of development [11-13].

*Allium cepa*, a popular condiment herb, is recognized for its various medicinal and pharmacological properties. This high-fiber perennial herb from the Amaryllidaceae family contains folic acid, vitamin B6, magnesium, calcium, potassium, phosphorus, and other vitamins and minerals. It possesses antibacterial, anticancer, antidiabetic, antioxidant, antiplatelet, antihypertensive, antidepressant, neuroprotective, anti-inflammatory, and antiparasitic effects [14,15]. This study aimed to find the natural mutated EGFR inhibitor to manage lung cancer.

## Methods

The crystal structure of the mutated EGFR protein with a co-crystallized ligand (PDB ID: 5CAO) was retrieved from the Protein Data Bank. The protein structure was preprocessed, minimized, and refined using the Discovery Studio Visualizer 2020. Preprocessing steps included removing crystallographic waters, assigning bond orders, and adding missing hydrogens.

### Compound library preparation

The bioactive compound library was created using data from the LOTUS database on *A. cepa* constituents. 251 distinct compounds were discovered, each with their own chemical structure and molecular properties. The compounds were downloaded in .sdf format and then processed for virtual screening (VS).

### Virtual Screening

High-throughput screening is the primary method for discovering novel lead compounds in drug development. It involves physically testing large chemical libraries against a specific biological target. VS is a complementary technique that uses computational techniques to evaluate large chemical repositories to identify molecules that complement structurally defined targets. The promising compounds are then experimentally validated [16,17]. Receptor-based methodologies, also known as structure-based techniques, seek to understand the interaction dynamics between a ligand and its protein to distinguish between high affinity and low affinity ligands.

These methods require a 3D target structure obtained via X-ray crystallography, NMR, or homology modeling. The availability of resolved protein structures has led to an increase in the use of receptor-based methods. Structure-based VS predicts the position and orientation of a ligand when it interacts with a protein [18]. PyRx 0.8 [19] was used in this study for VS (molecular docking) of *A. cepa* bioactive compounds with EGFR to determine the binding conformations with the lowest binding energies.

## Results

Findings from this study show that 100 (89.2%) samples the mutant EGFR protein employed in this study had a cocrystal ligand in its PDB structure. To validate the docking process, the cocrystal ligand was removed and re-docked to the EGFR. The results revealed that the docked co-crystal ligand binds in the identical EGFR binding area as seen in the co-crystal PDB structure (Figure 1). This consistency demonstrates the docking protocol's ability to accurately predict ligand binding orientation in the EGFR protein's binding pocket.

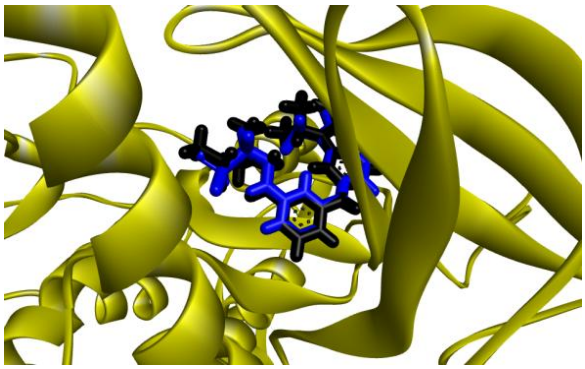


**Figure 1**: Representation of X-ray-bound cocrystal ligand (black) and re-docked cocrystal ligand (blue) in the EGFR protein.

The cocrystal ligand was used as a reference compound, exhibiting a binding energy of -8.6 kcal/mol with the EFGR protein. Compounds with binding energies more negative than the reference ligand were selected as hit compounds.

Consequently, 18 hit compounds were identified with binding energies greater than that of the cocrystal ligand, indicating stronger binding affinities for the EFGR protein (Table 1). Among these 18 compounds, the top five (LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270) were chosen for in-depth interaction analysis.

| S. No. | Compounds | Binding energy (kcal/mol) |
|---|---|---|
| 1. | LTS0258243 | -10.6 |
| 2. | LTS0042303 | -10.5 |
| 3. | LTS0058192 | -10.2 |
| 4. | LTS0104946 | -10.1 |
| 5. | LTS0145270 | -10.1 |
| 6. | LTS0143926 | -9.9 |
| 7. | LTS0200727 | -9.9 |
| 8. | LTS0084451 | -9.7 |
| 9. | LTS0030362 | -9.6 |
| 10. | LTS0161982 | -9.4 |
| 11. | LTS0221386 | -9.4 |
| 12. | LTS0253498 | -9.3 |
| 13. | LTS0125203 | -9.1 |
| 14. | LTS0108489 | -9.1 |
| 15. | LTS0189353 | -9.0 |
| 16. | LTS0196163 | -9.0 |
| 17. | LTS0267837 | -8.9 |
| 18. | LTS0204085 | -8.9 |
| 19. | Cocrystal ligand/inhibitor | -8.6 |

**Table 1:** List of top hit compounds having higher binding energies than the cocrystal ligand/inhibitor.

The interacting residues of EGFR protein with the top five compounds (LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270) and cocrystal ligand/inhibitor has been represented in figure 2 and table 2.

The structure and H-bonding residues of top five compounds (LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270) has been depicted in table 2.

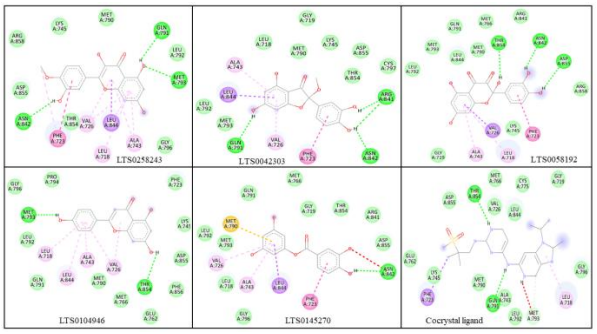


**Figure 2:** Interacting residues of EGFR protein with the top five compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 as well as the cocrystal ligand/inhibitor.

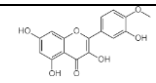
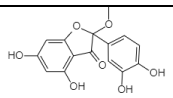
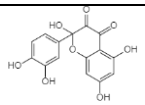
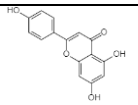
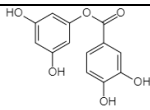
| Compounds | Structure | Interacting amino acids residues | H-bonding residues |
|---|---|---|---|
| LTS0258243 |  | Asn842, Asp855, Arg858, Lys745, Met790, Gln791, Leu792, Met793, Gly796, Ala743, Leu844, Leu718, Val726, Thr854, and Phe723 | Gln791, Met793, and Asn842 |
| LTS0042303 |  | Gln791, Met793, Leu792, Leu844, Ala743, Leu718, Met790, Gly719, Lys745, Asp855, Thr854, Cys797, Arg841, Asn842, Phe723, and Val726 | Arg841, Asn842, and Gln791 |
| LTS0058192 |  | Gly719, Leu792, Met793, Leu844, Gln791, Met790, Thr854, Met766, Arg841, Asn842, Asp855, Arg858, Phe723, Lys745, Val726, Leu718, and Ala743 | Thr854, Asn842, and Asp855 |
| LTS0104946 |  | Met793, Gly796, Pro794, Phe723, Lys745, Asp855, Phe856, Thr854, Glu762, Met766, Val726, Met790, Ala743, Leu844, Leu718, Gln791, and Leu792 | Thr854, and Met793 |
| LTS0145270 |  | Phe723, Leu844, Ala743, Gly796, Leu718, Val726, Met793, Leu792, Met790, Gln791, Met766, Gly719, Thr854, Arg841, Asp855, and Asn842 | Asn842 |

**Table 2:** Interacting residues as well as H-bonding residues of top five compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 with EGFR protein.

The top five compounds (LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270) had excellent molecular characteristics. The possibility that a substance is a natural product goes from -5 (less likely) to 5 (very likely). The top five compounds (LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270) had ratings of 1 or +1, indicating a strong preference for characteristics similar to those of natural products. Furthermore, Lipinski's Rule of Five revealed a comprehensive profile of these hits, measuring the drug's chemical structure to predict its

oral bioavailability, and they all satisfied the criteria with no violations (Table 3).

| Molecular properties and molecular descriptors | LTS0258243 | LTS0042303 | LTS0058192 | LTS0104946 | LTS0145270 |
|---|---|---|---|---|---|
| Total atom number | 35 | 34 | 33 | 30 | 29 |
| Heavy atom number | 23 | 22 | 23 | 20 | 19 |
| Bond count | 25 | 24 | 25 | 22 | 20 |
| Number of carbons | 16 | 15 | 15 | 15 | 13 |
| Minimal number of rings | 3 | 3 | 3 | 3 | 2 |
| Maximal number of rings | 4 | 4 | 4 | 4 | 2 |
| NP-likeness score | 1 | 1.02 | 1.04 | 1 | 1.01 |
| Alogp | 2.74 | 2.01 | 1.34 | 3.03 | 2.19 |
| Alogp2 | 7.53 | 4.06 | 1.8 | 9.16 | 4.81 |
| Apol | 41.7755 | 40.0155 | 39.4839 | 37.0779 | 34.3599 |
| Bpol | 17.9085 | 17.9085 | 14.7641 | 13.8061 | 13.8061 |
| Eccentric Connectivity Index Descriptor | 420 | 354 | 379 | 338 | 312 |
| Fmf Descriptor | 0.6957 | 0.6818 | 0.6957 | 0.8 | 0.7368 |
| Fsp3 | 0.0625 | 0.1333 | 0.0667 | 0 | 0 |
| Fragment Complexity Descriptor | 863.07 | 834.07 | 719.08 | 644.05 | 558.06 |
| Petitjean Number | 0.4545 | 0.5 | 0.5 | 0.5 | 0.5 |
| Lipinski's Rule (Failures) | 0 | 0 | 0 | 0 | 0 |
| WienerPathNumber | 1136 | 942 | 1072 | 788 | 746 |
| Xlogp | 3.338 | 1.426 | 0.95 | 3.668 | 2.061 |
| Zagreb Index | 124 | 122 | 128 | 108 | 96 |
| TopoPSA | 120.36 | 116.45 | 144.52 | 90.9 | 107.22 |

**Table 3:** Molecular properties and descriptors of top five compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270.

## Discussion

Cancer is a leading cause of death worldwide, with lung cancer being the most common and deadly type [20,21]. Despite various chemotherapeutic options, drug resistance in lung cancer remains a major challenge, requiring the development of new, effective treatments [22]. Plant-based medicines are widely used in many nations [23,24]. Considering that double-mutated EGFR is a key oncogenic protein in many lung cancer cases [6,25], this study focuses on screening of bioactive compounds from *A. cepa* against the 3D structure of double-mutated EGFR to identify potential strong-binding molecules. Among them, 18 compounds were found to strongly bind with the EGFR protein and have lower binding energy than the cocrystal ligand, with the top five hit compounds being LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270.

The hit compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 form several H-bonding with the EGFR amino acid residues. The Asn842 was the common H-bonding residue of EGFR with LTS0258243, LTS0042303, LTS0058192, and LTS0145270. In addition, LTS0104946 was H-bonded with Thr854, and Met793 residues of EGFR.

The Phe723, Lys745, Glu762, Asp855, Thr854, Met766, Val726, Leu844, Cys775, Gly719, Gly796, Leu718, Met793, Leu792, Ala743, Gln791, Met790, and Lys745 residues of EGFR were found to important in binding with cocrystal ligand. Interestingly, he hit compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 were also observed to interact with most of these EGFR residues.

In docking investigations, it is usually critical to seek a ligand that can bind effectively to the protein, with binding energy serving as a metric for better binding [26,27]. The free energy of binding is used to determine the strength of an interaction between a ligand and a protein complex [28-30]. The lowest binding energy is the result of the ligand's effective binding to the active site of the protein [31,32]. A higher (negative) binding energy indicates an efficient interaction between ligand and protein. Notably, the hit compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 had a robust interaction with the EGFR, with a high binding energy compared to the cocrystal ligand, indicating that these compounds might be potential EGFR inhibitors.

Lung cancer is among the primary causes of cancer-related mortality. Double-mutated EGFR is a major oncogenic protein in many lung cancer patients. This study screened the bioactive compounds in *A. cepa* against the double-mutated EGFR. The compounds LTS0258243, LTS0042303, LTS0058192, LTS0104946, and LTS0145270 were tightly bound to the EGFR and have a lower binding energy than the cocrystal ligand. In addition, these compounds exhibit strong drug-like characteristics and can be employed as EGFR inhibitors to manage lung cancer.

## Acknowledgement

I would like to express my sincere gratitude to the Faculty of Nursing at King Abdulaziz University, Jeddah, Saudi Arabia, for their invaluable support in the completion of this manuscript.

## Conflict of Interest

No conflict of interest.

## References

1. Release Notice - Canadian Cancer Statistics: A 2020 special report on lung cancer. Health Promotion and Chronic Disease Prevention in Canada, (2020); 40(9): 325.
2. Inamura K, Ninomiya H, Ishikawa Y, Matsubara O. Is the epidermal growth factor receptor status in lung cancers reflected in clinicopathologic features? Archives of pathology & laboratory medicine, (2010); 134(1): 66-72.
3. Ohsaki Y, Tanno S, Fujita Y, Toyoshima E, Fujiuchi S, et al. Epidermal growth factor receptor expression correlates with poor prognosis in non-small cell lung cancer patients with p53 overexpression. Oncology reports, (2000); 7(3): 603-610.
4. Scagliotti GV, Selvaggi G, Novello S, Hirsch FR. The biology of epidermal growth factor receptor in lung cancer. Clinical Cancer Research, (2004); 10(12): 4227-4232.
5. Veale D, Kerr N, Gibson G, Kelly P, Harris A. The relationship of quantitative epidermal growth factor receptor expression in non-small cell lung cancer to long term survival. British journal of cancer, (1993); 68(1): 162-165.
6. Khaddour K, Jonna S, Deneka A, Patel JD, Abazeed ME, et al. Targeting the epidermal growth factor receptor in EGFR-

mutated lung cancer: current and emerging therapies. Cancers, (2021); 13(13): 3164.
7. Agarwal SM, Pal D, Gupta M, Saini R. Insight into discovery of next generation reversible TMLR inhibitors targeting EGFR activating and drug resistant T790M mutants. Current Cancer Drug Targets, (2017); 17(7): 617-636.
8. Saini R, Fatima S, Agarwal SM. TMLRpred: A machine learning classification model to distinguish reversible EGFR double mutant inhibitors. Chemical Biology & Drug Design, (2020); 96(3): 921-930.
9. Das AP, Mathur P, Agarwal SM. Machine Learning, Molecular Docking, and Dynamics-Based Computational Identification of Potential Inhibitors against Lung Cancer. ACS omega, (2024); 9(4): 4528-4539.
10. Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. Jama, (2020); 323(9): 844-853.
11. Macalino SJY, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. Archives of pharmacal research, (2015); 38: 1686-1701.
12. Surabhi S, Singh B. Computer aided drug design: an overview. Journal of Drug delivery and Therapeutics, (2018); 8(5): 504-509.
13. Bharatam PV. Computer-aided drug design. Drug discovery and development: From targets and molecules to medicines, (2021); 137-210.
14. Chakraborty AJ, Uddin TM, Matin Zidan BR, Mitra S, Das R, et al. Allium cepa: a treasure of bioactive phytochemicals with prospective health benefits. Evidence-Based Complementary and Alternative Medicine, (2022); 2022(1): 4586318.
15. Marrelli M, Amodeo V, Statti G, Conforti F. Biological properties and bioactive components of Allium cepa L.: Focus on potential benefits in the treatment of obesity and related comorbidities. Molecules, (2018); 24(1): 119.
16. Lionta E, Spyrou G, K Vassilatis D, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Current topics in medicinal chemistry, (2014); 14(16): 1923-1938.
17. Shoichet BK. Virtual screening of chemical libraries. Nature, (2004); 432(7019): 862-865.
18. Ballante F, Kooistra AJ, Kampen S, de Graaf C, Carlsson J. Structure-based virtual screening for ligands of G protein–coupled receptors: what can molecular docking do for you? Pharmacological Reviews, (2021); 73(4): 1698-1736.
19. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Chemical biology: methods and protocols, (2015); 1263: 243-250.
20. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA: a cancer journal for clinicians, (2023); 73(1): 17-48.
21. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, (2021); 71(3): 209-249.
22. Yuan M, Huang L-L, Chen J-H, Wu J, Xu Q. The emerging treatment landscape of targeted therapy in non-small-cell lung cancer. Signal transduction and targeted therapy, (2019); 4(1): 61.
23. Atanasov AG, Waltenberger B, Pferschy-Wenzig E-M, Linder T, Wawrosch C, et al. Discovery and resupply of pharmacologically active plant-derived natural products: A review. Biotechnology advances, (2015); 33(8): 1582-1614.
24. Das AP, Agarwal SM. Recent advances in the area of plant-based anti-cancer drug discovery using computational approaches. Molecular Diversity, (2024); 28(2): 901-925.
25. Castaneda-Gonzalez JP, Chaves JJ, Parra-Medina R. Multiple mutations in the EGFR gene in lung cancer: a systematic review. Translational Lung Cancer Research, (2022); 11(10): 2148-2163.
26. Shaikh S, Aaqil H, Rizvi SMD, Shakil S, Abuzenadah AM, et al. Comparative inhibition study of compounds identified in the methanolic extract of *Apamarga Kshara* against Trichomonas vaginalis carbamate kinase (TvCK): an enzoinformatics approach. Interdisciplinary Sciences: Computational Life Sciences, (2016); 8: 357-365.
27. Alqahtani LS, Alkathiri AS, Alzahrani A, Alghamdi RM, Alamri WA, et al. Structure-Based Virtual Screening of Antiviral Compounds Targeting the Norovirus RdRp Protein. Advancements in Life Sciences, (2024); 11(2): 488-492.
28. Rafeeq MM, Helmi N, Sain ZM, Iqbal J, Alzahrani A, et al. Target-based virtual screening and molecular dynamics approach to identify potential antileishmanial agents through targeting UvrD-like helicase ATP-binding domain. Advancements in Life Sciences, (2024); 11(1): 237-245.
29. Alshehri MA, Asiri SA, Alzahrani A, Alazragi RS, Alqahtani LS, et al. Multitargeted inhibitory effect of Mitoxantrone 2HCl on cervical cancer cell cycle regulatory proteins: a multitargeted docking-based MM\GBSA and MD simulation study. Medical Oncology, (2023); 40(11): 337.
30. Alghamdi S, Baeissa HM, Azhar Kamal M, Rafeeq MM, Al Zahrani A, et al. Unveiling the multitargeted potency of Sodium Danshensu against cervical cancer: a multitargeted docking-based, structural fingerprinting and molecular dynamics simulation study. Journal of Biomolecular Structure and Dynamics, 2024; 42(16): 8268-8280.
31. Kamal MA, Baeissa HM, Hakeem IJ, Alazragi RS, Hazzazi MS, et al. Insights from the molecular docking analysis of EGFR antagonists. Bioinformation, (2023); 19(3): 260-265.
32. Alghamdi S. Molecular docking analysis of AGTR1 antagonists. Bioinformation, (2023); 19(3): 284-289.