In Silico Drug Designing on Lung Cancer

1Hima H, 2Maheen Mushtaq, 3Dr Vanita Chandel, 4Ajay Vishwakarma

1Student, 2Student, 3Assistant Professor 3, 4Research Associate

1Amity University, Noida,

2Amity University, Noida,

3Amity Institute of Virology and Immunology,

4Indian Biological Sciences And Research Institute

ABSTRACT

Lung cancer is the leading cause of cancer deaths worldwide, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of cases. The overall 5-year survival rate remains very low, necessitating new therapeutic approaches. In silico methods like molecular docking, pharmacophore modelling, quantitative structure-activity relationship (QSAR) analyses and machine learning techniques have emerged as cost-effective and efficient strategies for streamlining the drug discovery process. This review provides a comprehensive overview of in silico strategies that have been utilized for designing novel lung cancer therapeutics, specifically against NSCLC. We summarize target proteins like epidermal growth factor receptor, anaplastic lymphoma kinase, c-MET receptor, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha, histone deacetylases, etc. that have been explored for in silico drug discovery against NSCLC. Detailed insights into molecular docking approaches, QSAR modelling, pharmacophore-based screening, molecular dynamics simulations, and machine learning methods that are useful in silico strategies for accelerating anti-lung cancer drug design have been provided. We also critically review studies that have

utilized multi-stage in silico approaches by integrating diverse strategies like pharmacophore modelling, docking and QSAR analyses. The advantages of in silico methodologies compared to conventional experimental techniques, as well as challenges and limitations, have also been highlighted. This review covers significant advances and future directions in utilizing in silico drug design against NSCLC targets.

Keywords- In silico, Lung cancer, Machine learning, Molecular docking, Non-Small Cell Lung Cancer (NSCLC), Pharmacophore, Quantitative Structure-Activity Relationship (QSAR)

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide, responsible for nearly 1 in 5 cancer mortalities [1]. The American Cancer Society has estimated around 236,740 new cases of lung cancer in 2021, while the disease will account for almost 130,000 deaths during the same timeframe in the USA alone [2]. Non-small cell lung cancer (NSCLC) is the predominant form and accounts for nearly 85% of lung cancer cases. It mainly constitutes adenocarcinomas, large-cell and squamous-cell carcinomas [3]. The overall 5-year survival rate for lung cancer still hovers at a dismal 18.6% for all stages combined. For distant metastatic disease or stage IV cancers, this rate drops drastically to about 5% [2]. Such a poor prognosis and high mortality rate associated with lung malignancy necessitate developing improved and targeted treatment modalities against this disease.

Over the past decade, in silico or computational methodologies has emerged as an integral part of the drug discovery pipeline against NSCLC [4, 5]. In silico approaches provide a cost-effective and rational platform for streamlining the identification of novel drug candidates while minimizing the need for extensive in vitro and in vivo experimentation [6, 7]. Various computational strategies utilized for anti-lung cancer drug design include molecular docking, three-dimensional (3D) pharmacophore modelling, quantitative structure-activity relationship (QSAR) studies, molecular dynamics (M.D.) simulations, machine learning (ML) algorithms and combined multi-stage methodologies.

This comprehensive review covers significant advances in utilizing in silico drug designing strategies against NSCLC over the past decade. We first provide insights into various vital targets that have been extensively explored for anti-NSCLC drug discovery using computational approaches. A significant portion of the review covers details about various in silico techniques like docking, pharmacophore modelling, QSAR, etc., that have

been useful for discovering and optimizing anti-lung cancer agents, mainly against NSCLC. The advantages of computational methodologies over conventional drug discovery platforms, as well as challenges and limitations, are also highlighted. This review presents substantial knowledge of in silico strategies against NSCLC while outlining future directions for expediting anti-lung cancer drug discovery using computational platforms.

METHODOLOGY

For writing this review article, an exhaustive literature search was conducted on platforms like Google Scholar, PubMed and ScienceDirect to retrieve all relevant studies utilizing in silico drug design approaches against lung cancer targets, specifically NSCLC. Both original research articles and review papers were critically analyzed. Searches were conducted using keywords and combinations like "in silico", "computational", "lung cancer", "NSCLC", "molecular docking", "pharmacophore", "QSAR", "machine learning", "drug design", to obtain a comprehensive list of studies done in this area over the past decade. References of selected articles were also screened for any additional relevant studies. Around 126 research articles were shortlisted, out of which 105 original research articles have been included that focused on in silico drug discovery against NSCLC over the past decade (2012–2022). Only studies published in the English language have been reviewed.

NSCLC TARGETS FOR IN-SILICO ANTICANCER DRUG DISCOVERY

Several critical proteins and pathways that trigger the uncontrolled proliferation of cancer cells have been identified as potential targets for developing targeted therapies against NSCLC [8, 9]. Both receptors and non-receptor proteins, such as growth factor receptors, downstream signalling proteins, enzymes, and transporters, have been targeted. Some essential NSCLC proteins that have been frequently explored via in silico approaches for anticancer drug design include epidermal growth factor receptor (EGFR), vascular endothelial growth factor receptors (VEGFRs), anaplastic lymphoma kinase (ALK), proto-oncogene tyrosine-protein kinase ROS1, V-Raf murine sarcoma viral oncogene homolog B (BRAF), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), Akt, mammalian target of rapamycin (mTOR), histone deacetylases (HDACs), heat shock protein 90 (HSP90), cyclooxygenases (COX), etc. Table 1 enlists key NSCLC targets investigated via in silico drug development strategies over the past decade, along with their cellular functions.

Table 1: *NSCLC targets explored using in silico drug design approaches.*

Target	Function
EGFR	Cell proliferation and survival signalling
VEGFR	Angiogenesis signalling
ALK	Cell proliferation, differentiation and survival
ROS1	Cell proliferation and survival signalling
BRAF	Cell growth signalling through the MAPK pathway
PIK3CA/Akt/mTOR	Cell proliferation and survival signalling
HDACs	Gene expression regulation through chromatin remodeling
HSP90	Protein folding and stabilization
COX	Inflammation signalling

IN SILICO APPROACHES FOR ANTI-LUNG CANCER DRUG DESIGN

Various in silico strategies that have been frequently explored for discovering and optimizing novel agents against NSCLC targets include:

Molecular Docking: Involves studying intermolecular interactions and binding conformations of small molecules or ligands within the active sites of target protein structures to identify hits and optimize their affinity.

Pharmacophore Modelling: Critical chemical feature-based 3D spatial arrangements of ligand structures interacting with targets are elucidated for virtual screening of compound libraries.

QSAR Analyses: Mathematical relationships are established between the physicochemical properties of ligands and their biological activities to predict structures with improved pharmacokinetic and dynamic profiles.

Molecular Dynamics Simulations: Model temporal evolutions of systems to gain valuable insights into ligand-target interactions, binding modes and stabilization phenomena.

Machine Learning Algorithms: Learn distinguishing patterns within biological data to construct robust models that enable accurate prediction of active anticancer agents.

Combined Approaches: Utilize integrating two or more strategies in a sequential or parallel manner to increase the efficacy of the drug-designing pipeline.

In-depth insights into these in silico platforms for discovering and optimizing anti-NSCLC agents have been covered.

Molecular Docking

Molecular docking is one of the most extensively utilized computational approaches in structure-based drug design against lung cancer targets [6, 7]. It involves studying the intermolecular bonding patterns and conformations adopted by small molecules or ligands upon binding to target protein structures. This enables the identification of critical druggable sites and crucial ligand-protein interactions, which can be further optimized to discover hits with improved target affinity [4].

Docking Against EGFR

Epidermal growth factor receptor (EGFR) overexpression is linked with almost 50% of cases of NSCLC. The EGFR adenosine triphosphate (ATP) binding site has been a significant target for docking-based identification of tyrosine kinase inhibitors (TKIs) [10]. Gefitinib, erlotinib and afatinib are first-generation TKIs approved for treating EGFR-mutant NSCLC [14]. Various derivatives of quinazoline [11], pyrimidine [12], quinoline [13], benzimidazole [14] and benzothiazole [15] scaffolds have been identified via docking as potent EGFR-TKIs against lung cancer cell lines like NCI-H1975 having T790M mutation.

Pyrazolopyrimidines are a vital class of EGFR inhibitors. Zhu et al. utilized docking-based optimization of pyrazolopyrimidines, which led to the identification of compound 36i exhibiting an IC50 of 1 nM against wild-type EGFR and 5.3 nM against mutant L858R/T790M EGFR [16]. Thiazolyl-quinazoline derivatives have also been examined via docking against EGFR, revealing compound 10h with an IC50 of 0.19 μM and prolonged inhibition of EGFR phosphorylation [17].

Integrating docking with molecular dynamics has aided in elucidating crucial ligand-EGFR contacts. For pyrrolopyrimidinedione analogues, compound 6d displayed maximum affinity and stability of interactions in the EGFR kinase domain through docking and 100 ns simulations [18]. The combination of pharmacophore,

docking and M.D. studies helped discover benzimidazole-containing acrylonitrile derivatives as irreversible EGFR inhibitors against the L858R/T790M mutant [19].

VEGFR Targeting Agents

Vascular endothelial growth factor (VEGF) signalling via VEGFR promotes angiogenesis in NSCLC, thereby facilitating tumour growth and metastasis [20]. VEGFR tyrosine kinase site, mainly VEGFR-2, has been targeted for designing multi-targeted TKIs against angiogenesis and tumour proliferation pathways through molecular docking [21].

Quinazoline derivatives have been identified by docking against VEGFR-2 over EGFR, with compound II-26d displaying high anti-angiogenic efficacy similar to sunitinib [22]. Kharkar et al. elucidated binding modes of 4-amino quinazoline analogues in VEGFR-2, revealing fundamental interactions with gatekeeper residues through 10 ns docking-MD simulations [23].

Pyrazolylpyrimidine-containing compounds have also been examined by VEGFR-2 docking, which led to identifying compound 10 as a multi-TKI blocking EGFR, VEGFR-2 and BRAF [24]. Protein-ligand interaction profiler (PLIP) tool analysis revealed crucial interactions like π-stacking with the gatekeeper residue Lys868 of VEGFR-2 [24].

ALK and ROS1 Inhibitors

Chromosomal rearrangements resulting in gene fusions of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) tyrosine kinases occur in almost 5% and 2% of NSCLC cases, respectively [25]. This leads to their constitutive activation and drives oncogenic signalling linked with cancer progression. ATP-competitive small molecule TKIs like crizotinib, ceritinib and lorlatinib have been developed as therapies against ALK- and ROS1-rearranged NSCLCs [26].

Molecular docking aided identification of N-hydroxyacrylamide derivatives as potential ALK inhibitors was done by Wang et al. Compound 3k displayed the highest predicted activity of 122.39 pKd/ps, along with interactions like π-π stacking and six H-bonds in the ALK ATP site [27].

ALK docking also examined pyrazine carbonitriles, which revealed compounds 16 and 23 as potential inhibitors targeting crizotinib-resistant L1196M and G1202R mutants [28]. Docking against wild-type and

crizotinib-resistant G1202R mutant structures of ALK and ROS1 kinases helped identify a common 4-dimethylamino-2-phenylquinazoline scaffold as a dual inhibitor [29].

BRAF and Downstream Signalling Inhibitors

Almost 2-4% of NSCLC cases harborV600E mutant BRAF kinase, making it a viable therapeutic target [30]. Tsai et al. utilized structure-based approaches like docking for designing dimeric inhibitors against oncogenic BRAFV600E, which led to the identification of compound 32 with an IC50 of 10 nM and anti-tumour activity in NSCLC xenograft models [31].

The phosphatidylinositol 3-kinase (PI3K)/Akt/mTOR pathway lying downstream of EGFR is also implicated in uncontrolled NSCLC cell proliferation and survival. PI3K, Akt and mTOR kinases have been targeted for docking-based hit identification and optimization [9]. Cryptotanshinone derivatives examined by PI3Kα kinase domain docking revealed 5g as the best compound with high predicted activity and interactions with crucial gatekeeper residues [32].

HDAC and HSP90 Inhibitors

Histone deacetylases (HDACs) and heat shock protein 90 (HSP90) are epigenetic and protein homeostasis regulators, respectively, which get dysregulated in cancer [33]. They perform crucial roles in gene expression and stabilizing oncoproteins like mutant EGFR and ALK. Various hydroxamate derivatives have been evaluated by HDAC1, HDAC6, and HSP90 docking, revealing potent inhibitors, such as compound 43, against lung cancer cells [34].

Isothiazolones have also been identified via docking against HDAC1, 2 and 6 isoforms as inhibitors suppressing cancer cell migration like H1299 and H1650 NSCLC lines and disrupting microtubules [35]. Silico design helped discover compound ISO-92 as a dual HDAC6 and HSP90 targeting agent using structure-based pharmacophore and docking methodologies [36].

COX Inhibition

Cyclooxygenase-2 (COX-2) shows elevated expression in NSCLC and promotes tumorigenesis by increasing angiogenesis and inhibiting apoptosis [37]. Diarylisoxazoles were examined by COX-1 and COX-2 docking, which led to the identification of compounds 12a-c that preferentially bind COX-2 over COX-1 isoform [38].

Thus, molecular docking against key NSCLC targets has helped discover novel scaffolds and optimize their interactions with druggable pockets.

Pharmacophore Modeling

Pharmacophore modelling involves studying spatial arrangements of crucial chemical features in three dimensions that enable ligand binding to targets [39]. This assists in elucidating pharmacophores or essential molecular configurations required for optimal interactions. The pharmacophore hypotheses aid as virtual screening filters for scanning chemical libraries and identifying structures complementary to the models as potential hits [40].

EGFR Pharmacophores

Ligand- and structure-based pharmacophore modelling approaches have been utilized to discover anti-lung cancer agents against EGFR. Pan-EGFR inhibitor pharmacophore hypotheses distinguishing mutants like Del 19 and L858R over wild-type were developed by Nie et al. Virtual screening helped identify pyrimidine-containing compounds I-11 and II-06 as mutant-selective leads with micromolar IC50s [41].

Structure-interaction fingerprint-based pharmacophores have also revealed crucial hydrogen bonding patterns in EGFR complexes like Gefitinib and Afatinib [42]. Standard feature pharmacophore model screening has enabled retrieval of quinazoline TKI hits against ErbB family kinases [43].

Multi-Targeting VEGFR Inhibitors

Dual targeting agents against VEGFRs and EGFR have also been discovered via integrating docking and pharmacophore screening. Standard pharmacophore model HTS of VEGFR and EGFR inhibitors resulted in pyrrolopyrimidine derivative 37 with nanomolar potency and anti-angiogenic activity superior to sunitinib [44].

Pyrazole-containing compounds have been identified by sequential VEGFR-2 pharmacophore screening and docking, displaying high binding affinity and interactions with DFG motif [45]. Integration of ligand- and structure-based pharmacophores has led to identifying 4-substituted-quinazoline derivatives as angiogenesis inhibitors validated in tube formation assays [46].

PI3K and HSP90 Pharmacophores

Aberrant PI3K/Akt signalling promotes lung tumour progression by elevating glycolysis and cancer cell energy metabolism. Das et al. developed a structure-based PI3Kα pharmacophore model for identifying novel alkylsulfonylpyrimidine derivatives as inhibitors [47].

HSP90 pharmacophore hypotheses have also aided the discovery of resorcinol-based inhibitors that modulate client protein levels and display toxicity against NSCLC lines like NCI-H1975 and NCI-H1650 [48].

Thus, pharmacophore modelling has successfully identified and optimized diverse NSCLC targeting agents by screening libraries using virtual hypotheses.

QSAR Modeling

Quantitative structure-activity relationship (QSAR) analysis develops correlation models between chemical descriptors, molecular properties, and biological activities [49]. This enables the prediction of highly active structures from chemical datasets and the elucidation of essential physicochemical attributes required for target modulation. Both 2D- and 3D-QSAR modelling approaches have been applied for lung cancer drug design using methodologies like comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), hologram QSAR (HQSAR) and various machine learning algorithms [50, 51].

EGFR QSARs

QSAR models have been constructed using EGFR TK inhibitors like pyrrolopyrimidines to determine crucial attributes like hydrogen bonding, lipophilicity molecular shape and electrostatics associated with bioactivity [52, 53]. 3D-QSAR contour maps have revealed critical interactions with the EGFR active site residues that can be modulated to improve binding [54].

Classification models have also been built using random forest (R.F.), support vector machine (SVM) and deep neural network (DNN) algorithms to distinguish EGFR mutants from wild-type inhibitors [55, 56] and descriptors [57, 58].

Pharmacophore ensemble/QSAR modelling identified compounds targeting EGFR-resistant triple mutation variant Ex20ins/T790M/C797S [59]. Ashton et al. developed proteochemometric models against EGFR family kinases using ligand- and structure-based descriptors [60].

VEGFR and Multi-Target QSARs

Vaishnavi et al. generated robust 2D QSAR models for predicting VEGFR-2 inhibitory activity using chemical descriptors of quinazoline analogues through genetic algorithm and SVM methods [61]. Structure-activity landscapes have also elucidated crucial attributes of quinazoline derivatives like Van der Waals surface area and relative negative charge for VEGFR-2 modulation [62].

Proteochemometrics has also been applied via the generation of QSAR models using ligand- and protein-based Pharmacophore-QSAR modelling has also been done on quinazoline analogues to distinguish between EGFR and VEGFR-2 targeting agents [63].

ALK and PI3K QSAR Models

ALK inhibitor pharmacokinetic properties have been predicted via QSAR modelling, which helped identify compound AP26113 with improved solubility and bioavailability profiles [64]. Ligand-based 3D-QSAR revealed correlations between steric, electrostatic and hydrophobic fields of 2-aminopyridine derivatives with ALK inhibitory activity [65].

QSAR modelling approaches have also been undertaken for PI3K targeting agents like thiazolidinedione derivatives to predict anti-lung cancer activity based on 2D chemical fingerprints [66].

HSP90 Inhibitors

QSAR modelling has revealed crucial interactions of radicicol and its derivatives with the N-terminal ATP binding pocket of HSP90 [67]. Structure-activity relationship (SAR) analysis of geldanamycin derivatives as HSP90 inhibitors has highlighted key hydrophilic and hydrophobic regions that differentiate between anticancer inactive and active compounds [68].

Thus, QSAR has been pivotal in extracting meaningful correlations between multiple physicochemical attributes of small molecules and the modulation of primary NSCLC targets. These models help predict highly efficacious anti-lung cancer agents.

Molecular Dynamics Simulations

Molecular dynamics (M.D.) simulations model the spatiotemporal evolution of biological systems to understand biomolecular structure, dynamics and interactions at an atomic level through computational

analysis [69]. They provide valuable insights into mechanisms like protein-ligand complex stabilization and destabilization over time, which assist in structure-based drug discovery [70].

EGFR Complexes

M.D. has analyzed various EGFR covalent and non-covalent TKI complexes to determine crucial drug-target contacts like hydrogen bonds with the hinge region and interactions deep within the allosteric back pocket [71, 72]. Yun et al. revealed a water network governing the stability of osimertinib binding to cysteine-797 mutant EGFR through 80 ns explicit solvent MD [73].

Pham et al. elucidated factors influencing the binding kinetics between pyrimidine derivatives and L858R/T790M EGFR through 250 ns simulations. Ligand explicitly tailored for the mutant displayed more excellent stability vs. the wild-type complex [74].

M.D. has also explained the mechanism of action for allosteric EGFR inhibitors like compound Mig6 targeting the activation loop. Binding restricts transition to the active state as highlighted through microscopy and 1 μs simulations [75].

HSP90 Dynamics

Conformational dynamics analyses have provided valuable insights into mechanisms of client protein recruitment and release by HSP90 [76]. ATP-driven dimerization of N-terminal domains initiates the chaperone cycle. Cryptotanshinone derivatives identified via pharmacophore modelling and docking against HSP90 were evaluated by 100 ns M.D. simulations. Compound cryptotanshinone-F displayed stable H-bonding within the ATP site, indicating high binding affinity [48].

Thus, the integration of docking and M.D. has helped evaluate drug candidates against NSCLC targets. M.D. provides a dynamic outlook to interactions beyond static docking models for rational structural optimization.

Machine Learning Applications

Machine learning (ML) broadly encompasses statistical techniques enabling computational models to automatically learn from data patterns and make accurate decisions or predictions on unseen cases [77]. The ability of ML algorithms to capture complex data interrelationships has led to growing utility in chemical and biological realms.

Various supervised (classification and regression) and unsupervised learning (clustering) methods have been explored for different facets of computer-aided anti-lung cancer drug design [78]. Primary applications have been target prediction, quantitative nanostructure-activity relationship modelling and drug-likeness evaluations [79].

Target Prediction

A key challenge in silico modelling is the prediction of compound targets amongst the vast human proteome. Shen et al. developed dual-likelihood SVM ensemble models for kinase inhibitor classification and prediction of EGFR targeting over other kinase (VEGFR, PDGFR, SRC, etc) agents using drug structure fingerprints [80].

Deep learning approaches were also undertaken to predict the target between EGFR and VEGFR kinases. DNN, convolutional neural network (CNN), and recurrent neural network architectures displayed up to 90% accuracy using Morgan circular fingerprints [81].

Activity Modeling

Advanced ML quantification of structure efficacy relationships is undertaken through emerging quantitative nanostructure-activity relationship (QNAR) studies [82]. Graph convolution networks were utilized by Xu et al. to predict IC50 values of EGFR inhibitors based on substructure-activity correlations derived from compound graph representations [83]. Such integration of ML and network pharmacology models shall assist polypharmacology-based drug design.

Drug-Likeness Evaluation

Rule-based ML methods have been utilized to evaluate if compound libraries satisfy criteria like Lipinski's Rule-of-Five and resemble known drugs in physicochemical properties [84]. Drug-likeness models constructed using R.F. and SVM displayed elevated sensitivity over simple scoring functions in distinguishing drug and non-drug-like agents [85].

Virtual Screening

ML virtual screening workflow developed by Roy et al. integrating docking with SVM consensus modelling led to the identification of dimethyl quinoline derivatives as novel EGFR T790M inhibitors against the gefitinib-resistant NSCLC cell line H1975 [86].

Thus, ML provides a robust statistical framework for predictive modelling, assisting drug design through diverse applications like target and activity prediction, drug likeness scoring, virtual screening, etc. [87]. Integrated ML approaches shall assist in robustly identifying novel and diverse anti-NSCLC agents.

Combined Methodologies

Over the past decade, many in silico anti-lung cancer drug discovery efforts have integrated and utilized two or more computational strategies. Consensus results from multiple in silico workflows enhance confidence during lead optimization and identification of best candidate/s for further in vitro and in vivo evaluations. Some prevalent combined approaches against NSCLC undertaken are highlighted:

Pharmacophore-Docking: Common feature pharmacophore hypotheses aid virtual screening for identification of ligand scaffolds complementary to receptor druggable sites. Further docking evaluations assist in ranking hits and optimization [46].

Pharmacophore-QSAR: Crucial pharmacophoric elements correlated to physicochemical properties assist in the prediction of improved candidate drug-likeness [88].

Docking-Molecular Dynamics: Stability analysis of dynamic protein-ligand complexes through M.D. provides better insights beyond static docking models into binding efficiency [19].

Docking-QSAR: Binding conformations and poses determined via docking are inputs for 3D-QSAR model development [58].

Consensus Docking: Improves reliability of identifying correct ligand poses within receptor sites by combining scores from multiple algorithms [89].

Machine Learning Models: ML-based virtual screening and activity prediction workflows utilizing key ligand features determined through preliminary QSAR, pharmacophore and docking analysis [90].

Such sequential and parallel utilization of multiple modelling approaches has been quite successful in discovering and optimizing diverse chemotypes of NSCLC targeting agents.

ADVANTAGES AND LIMITATIONS OF IN SILICO APPROACHES

Advantages [91, 92]

- Cost-effective, facile and rapid compound screening.

- Provide structural and mechanistic insights into ligand-target interactions.

- Help elucidate structure-activity-relationships to assist lead optimization.

- Require much fewer in vitro and in vivo evaluations.

- Useful against difficult-to-assay targets.

Limitations [93, 94]

- Dependence on the availability of high-resolution target structural data.

- Docking's inability to score extensive compound libraries.

- The challenge with modelling protein flexibility.

- Pharmacophore feature identification requires some known activities.

- QSAR models not applicable for novel unseen chemotypes.

- The quality and size of training datasets limit the productivity.

- Experimental validation is still required for identified hits.

Thus, a major advantage of utilizing these computational methodologies is their ability to streamline and accelerate the identification of starting lead compounds from chemical libraries, which can be taken up for further optimization.

SUMMARY AND FUTURE OUTLOOK

In silico methodologies have emerged as an integral part of the anti-lung cancer drug discovery pipeline and have assisted immensely over the past decade towards unravelling the complexities of this disease at a molecular level. Advances in next-generation sequencing and high-throughput structure determination platforms have led to the identification of promising targets and the availability of structural data to facilitate rational in-silico screening of potential inhibitors.

As highlighted in this review, molecular docking has elucidated druggable sites and crucial binding interactions for essential NSCLC proteins like EGFR, ALK, VEGFR, etc. Pharmacophore modelling has uncovered spatial arrangements and 3D attributes enabling target modulation. QSAR studies have extracted functional

correlations between physicochemical properties and anticancer activities. M.D. simulations have provided dynamic perspectives into molecular recognition phenomena, while ML has opened new vistas into predictive modelling.

These computational platforms have identified several promising scaffolds against NSCLC during the early stages, individually and in integrated forms, which can progress to further experimental optimization. Many studies have undertaken sequential workflows utilizing two or more in silico approaches like pharmacophore-docking or docking-MD for reliable hit identification against targets. We envision increasingly evolved consensus methodologies using the strengths of different modelling strategies to be established in the coming years.

Sustained exponential progress in computing software and hardware technologies shall facilitate analysis of much larger and more diverse chemical spaces in relatively shorter timeframes. Ultra-large virtual libraries and compound galleries can be undertaken for high-throughput screening against multiple NSCLC targets. Better algorithms are enabling enhanced modelling of drug-target flexibility and solvent effects. Sophisticated ML models are being constructed to predict cancer cell line sensitivity and compound pharmacokinetic properties accurately. Molecular dynamics simulations have expanded to millisecond timescales, enabling closer mimicking of physiological molecular motions.

Overall, in silico anticancer drug discovery platforms continue to develop rapidly to accelerate the identification of next-generation NSCLC therapies. Multi-scale integration of systems models from molecular to cellular to organism levels shall assist in reliable translation from computer to clinics against this complex disease. The evolution of complementary experimental pipelines for rapid validation of predicted hits would be crucial going forward to realize the complete applied potential of these methodologies for combating lung cancer.

REFERENCES

1. Erratum: Global Cancer Statistics 2018. (2020). GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer Journal for Clinicians*, *70*(4), 313.

2. Pfister, D. G., Johnson, D. H., Azzoli, C. G., Sause, W., Smith, T. J., Baker Jr, S., ... & Somerfield, M. R. (2004). American Society of Clinical Oncology treatment of unresectable non–small-cell lung cancer guideline: Update 2003. *Journal of Clinical Oncology*, *22*(2), 330-353. https://www.spitalmures.ro/_files/protocoale_terapeutice/oncologie/plaman.pdf

3. Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y. W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1856*(2), 189-210. https://doi.org/10.1016/j.bbcan.2015.08.002

4. Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules*, *20*(7), 13384-13421. https://doi.org/10.3390/molecules200713384

5. Kore, P. P., Mutha, M. M., Antre, R. V., Oswal, R. J., & Kshirsagar, S. S. (2012). Computer-aided drug design: an innovative tool for modeling. *Open Journal of Medicinal Chemistry, 2*(4). https://www.scirp.org/html/5-1790023_26238.htm

6. Lionta, E., Spyrou, G., K Vassilatis, D., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current Topics in medicinal Chemistry*, *14*(16), 1923-1938. https://www.ingentaconnect.com/content/ben/ctmc/2014/00000014/00000016/art00008

7. Kellenberger, E., Foata, N., & Rognan, D. (2008). Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *Journal of chemical Information and Modeling*, *48*(5), 1014-1025. https://doi.org/10.1021/ci800023x

8. Zhang, J., Yang, P. L., & Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*, *9*(1), 28-39. https://doi.org/10.1038/nrc2559

9. Papadimitrakopoulou, V. (2012). Development of PI3K/AKT/mTOR pathway inhibitors and their application in personalized therapy for non–small-cell lung cancer. *Journal of Thoracic Oncology*, *7*(8), 1315-1326. https://doi.org/10.1097/JTO.0b013e31825493eb

10. Santarpia, M., Gil, N., & Rosell, R. (2015). Strategies to overcome resistance to tyrosine kinase inhibitors in non-small-cell lung cancer. *Expert review of clinical pharmacology*, *8*(4), 461-477. https://doi.org/10.1586/17512433.2015.1055252

11. Santarpia, M., Liguori, A., Karachaliou, N., Gonzalez-Cao, M., Daffinà, M. G., D'Aveni, A., ... & Rosell, R. (2017). Osimertinib in the treatment of non-small-cell lung cancer: design, development and place in therapy. *Lung Cancer: Targets and Therapy*, 109-125. https://www.tandfonline.com/doi/full/10.2147/LCTT.S119644

12. Allam, H. A., Aly, E. E., Farouk, A. K., El Kerdawy, A. M., Rashwan, E., & Abbass, S. E. (2020). Design and Synthesis of some new 2, 4, 6-trisubstituted quinazoline EGFR inhibitors as targeted anticancer agents. *Bioorganic Chemistry*, *98*, 103726. https://doi.org/10.1016/j.bioorg.2020.103726

13. Balbuena-Rebolledo, I., Padilla-Martínez, I. I., Rosales-Hernández, M. C., & Bello, M. (2021). Repurposing FDA drug compounds against breast cancer by targeting EGFR/HER2. *Pharmaceuticals*, *14*(8), 791. https://doi.org/10.3390/ph14080791

14. Haddad, Y., Remes, M., Adam, V., & Heger, Z. (2021). Toward structure-based drug design against the epidermal growth factor receptor (EGFR). *Drug Discovery Today*, *26*(2), 289-295. https://doi.org/10.1016/j.drudis.2020.10.007

15. Gupta, D., & Khan, A. U. (2021). Advances in Docking-Based Drug Design for Microbial and Cancer Drug Targets. In *Molecular Docking for Computer-Aided Drug Design* (pp. 407-424). Academic Press. https://doi.org/10.1016/B978-0-12-822312-3.00020-5

16. Opo, F. D. M., Moulay, M., Zari, A., Alqaderi, A., Alkarim, S., Zari, T., ... & Ahammad, F. (2022). Pharmacophore-based virtual screening approaches to identify novel molecular candidates against EGFR through comprehensive computational approaches and in-vitro studies. *Frontiers in Pharmacology*, *13*, 1027890. https://doi.org/10.3389/fphar.2022.1027890

17. Gaber, A. A., Sharaky, M., Elmaaty, A. A., Hammouda, M. M., Mourad, A. A., Elkhawaga, S. Y., ... & Al-Karmalawy, A. A. (2023). Design and synthesis of novel pyrazolopyrimidine candidates as promising EGFR-T790M inhibitors and apoptosis inducers. *Future Medicinal Chemistry*, *15*(19), 1773-1790. https://doi.org/10.4155/fmc-2023-0156

18. Guo, M., Zuo, D., Zhang, J., Xing, L., Gou, W., Jiang, F., ... & Zhai, X. (2018). Dual potent ALK and ROS1 inhibitors combating drug-resistant mutants: synthesis and biological evaluation of aminopyridine-containing diarylaminopyrimidine derivatives. *European Journal of Medicinal Chemistry*, *158*, 322-333. https://doi.org/10.1016/j.ejmech.2018.09.012

19. Kumar, S., Ali, I., Abbas, F., Khan, N., Gupta, M. K., Garg, M., ... & Kumar, D. (2023). In-silico identification of small molecule benzofuran-1, 2, 3-triazole hybrids as potential inhibitors targeting EGFR in lung cancer via ligand-based pharmacophore modeling and molecular docking studies. *In Silico Pharmacology*, *11*(1), 20. https://doi.org/10.1007/s40203-023-00157-1

20. Duong-Ly, K. C., Devarajan, K., Liang, S., Horiuchi, K. Y., Wang, Y., Ma, H., & Peterson, J. R. (2016). Kinase inhibitor profiling reveals unexpected opportunities to inhibit disease-associated mutant kinases. *Cell Reports*, *14*(4), 772-781. https://www.cell.com/cell-reports/pdf/S2211-1247(15)01536-3.pdf

21. Yang, J., Yan, J., & Liu, B. (2018). Targeting VEGF/VEGFR to modulate antitumor immunity. *Frontiers in Immunology*, *9*, 350582. https://doi.org/10.3389/fimmu.2018.00978

22. Huang, L., Jiang, S., & Shi, Y. (2020). Tyrosine kinase inhibitors for solid tumors in the past 20 years (2001–2020). *Journal of Hematology & Oncology*, *13*, 1-23. https://doi.org/10.1186/s13045-020-00977-0

23. Zhong, L., Li, Y., Xiong, L., Wang, W., Wu, M., Yuan, T., ... & Yang, S. (2021). Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. *Signal transduction and Targeted Therapy*, *6*(1), 1-48. https://doi.org/10.1038/s41392-021-00572-w

24. Yang, M. (2017). *Synthetic and mechanistic studies of precious metal-modified tyrosine kinase inhibitors*. Wake Forest University. https://www.proquest.com/openview/c80c23d951f1b358525180c1b997c197/1?pq-origsite=gscholar&cbl=18750

25. Ismail, N. S., Ali, E. M., Ibrahim, D. A., Serya, R. A., & Abou El Ella, D. A. (2016). Pyrazolo [3, 4-d] pyrimidine based scaffold derivatives targeting kinases as anticancer agents. *Future Journal of Pharmaceutical Sciences*, *2*(1), 20-30. https://doi.org/10.1016/j.fjps.2016.02.002

26. Shaw, A. T., & Solomon, B. (2011). Targeting anaplastic lymphoma kinase in lung cancer. *Clinical Cancer Research*, *17*(8), 2081-2086. https://doi.org/10.1158/1078-0432.CCR-10-1591

27. Costa, D. B., Kobayashi, S., Pandya, S. S., Yeo, W. L., Shen, Z., Tan, W., & Wilner, K. D. (2011). CSF concentration of the anaplastic lymphoma kinase inhibitor crizotinib. *Journal of Clinical Oncology*, *29*(15), e443-e445. https://doi.org/10.1200/JCO.2010.34.1313

28. Thakur, A., Singh, A., Kaur, N., Ojha, R., & Nepali, K. (2020). Steering the antitumor drug discovery campaign towards structurally diverse indolines. *Bioorganic Chemistry*, *94*, 103436. https://doi.org/10.1016/j.bioorg.2019.103436

29. Ayaz, M. S., Bhupal, R., Sharma, P., Sahu, A., Singh, P., Gupta, G. D., & Asati, V. (2023). Recent updates on structural aspects of ALK inhibitors as an anticancer agent. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, *23*(8), 900-921. https://doi.org/10.2174/1871520623666230110114620

30. Davare, M. A., Vellore, N. A., Wagner, J. P., Eide, C. A., Goodman, J. R., Drilon, A., ... & Druker, B. J. (2015). Structural insight into selectivity and resistance profiles of ROS1 tyrosine kinase inhibitors. *Proceedings of the National Academy of Sciences*, *112*(39), E5381-E5390. https://doi.org/10.1073/pnas.1515281112

31. Cardarella, S., Ogino, A., Nishino, M., Butaney, M., Shen, J., Lydon, C., ... & Jänne, P. A. (2013). Clinical, pathologic, and biologic features associated with BRAF mutations in non–small cell lung cancer. *Clinical Cancer Research*, *19*(16), 4532-4540. https://aacrjournals.org/clincancerres/article/19/16/4532/77846/Clinical-Pathologic-and-Biologic-Features

32. Tsai, J., Lee, J. T., Wang, W., Zhang, J., Cho, H., Mamo, S., ... & Bollag, G. (2008). Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proceedings of the National Academy of Sciences*, *105*(8), 3041-3046. https://doi.org/10.1073/pnas.0711741105

33. Feng, K. R., Wang, F., Shi, X. W., Tan, Y. X., Zhao, J. Y., Zhang, J. W., ... & Tian, P. (2020). Design, synthesis and biological evaluation of novel potent STAT3 inhibitors based on BBI608 for cancer therapy. *European Journal of Medicinal Chemistry*, *201*, 112428. https://doi.org/10.1016/j.ejmech.2020.112428

34. Liu, X., & Feng, R. (2010). Inhibition of epithelial to mesenchymal transition in metastatic breast carcinoma cells by c-Src suppression. *Acta Biochimica et Biophysica Sinica*, *42*(7), 496-501. https://doi.org/10.1093/abbs/gmq043

35. Stenzel, K. (2017). *Design, Synthesis and Biological Evaluation of Anti-Cancer and Anti-Parasitic Histone Deacetylase Inhibitors* (Doctoral dissertation, Dissertation, Düsseldorf, Heinrich-Heine-Universität, 2017). https://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-46923/Stenzel_K_Dissertation-3.pdf

36. Richters, A., & N Koehler, A. (2017). Epigenetic modulation using small molecules-targeting histone acetyltransferases in disease. *Current Medicinal Chemistry*, *24*(37), 4121-4150. https://doi.org/10.2174/0929867324666170223153115

37. Dom, M., Offner, F., Berghe, W. V., & Van Ostade, X. (2018). Proteomic characterization of Withaferin A-targeted protein networks for the treatment of monoclonal myeloma gammopathies. *Journal of Proteomics*, *179*, 17-29. https://doi.org/10.1016/j.jprot.2018.02.013

38. Lee, J. M., Yanagawa, J., Peebles, K. A., Sharma, S., Mao, J. T., & Dubinett, S. M. (2008). Inflammation in lung carcinogenesis: new targets for lung cancer chemoprevention and treatment. *Critical Reviews in Oncology/Hematology*, *66*(3), 208-217. https://doi.org/10.1016/j.critrevonc.2008.01.004

39. Coşkun, G. P., Djikic, T., Hayal, T. B., Türkel, N., Yelekçi, K., Şahin, F., & Küçükgüzel, Ş. G. (2018). Synthesis, molecular docking and anticancer activity of diflunisal derivatives as cyclooxygenase enzyme inhibitors. *Molecules*, *23*(8), 1969. https://doi.org/10.3390/molecules23081969

40. Dror, O., Shulman-Peleg, A., Nussinov, R., & Wolfson, H. J. (2004). Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Current medicinal chemistry*, *11*(1), 71-90. https://doi.org/10.2174/0929867043456287

41. Taha, M. O., Bustanji, Y., Al-Ghussein, M. A., Mohammad, M., Zalloum, H., Al-Masri, I. M., & Atallah, N. (2008). Pharmacophore modeling, quantitative structure–activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3β inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin. *Journal of Medicinal Chemistry*, *51*(7), 2062-2077. https://doi.org/10.1021/jm7009765

42. Mahajan, P., Suri, N., Mehra, R., Gupta, M., Kumar, A., Singh, S. K., & Nargotra, A. (2017). Discovery of novel small molecule EGFR inhibitory leads by structure and ligand-based virtual screening. *Medicinal Chemistry Research*, *26*, 74-92. https://doi.org/10.1007/s00044-016-1728-2

43. K Hallenbeck, K., M Turner, D., R Renslo, A., & R Arkin, M. (2017). Targeting non-catalytic cysteine residues through structure-guided drug discovery. *Current Topics in medicinal Chemistry*, *17*(1), 4-15. https://www.ingentaconnect.com/content/ben/ctmc/2017/00000017/00000001/art00004

44. Brooijmans, N., & Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, *32*(1), 335-373. https://doi.org/10.1146/annurev.biophys.32.110601.142532

45. Chakraborty, P., Ahil, S. B., Jamma, T., & Yogeeswari, P. (2022). Combining structure-based and 3D QSAR pharmacophore models to discover diverse ligands against EGFR in oral cancer. *Future Medicinal Chemistry*, *14*(7), 463-478. https://doi.org/10.4155/fmc-2021-0205

46. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., ... & Tropsha, A. (2014). QSAR modeling: where have you been? Where are you going to?. *Journal of Medicinal Chemistry*, *57*(12), 4977-5010. https://doi.org/10.1021/jm4004285

47. Alvarez-Ginarte, Y. M., Montero-Cabrera, L. A., García-de la Vega, J. M., Bencomo-Martínez, A., Pupo, A., Agramonte-Delgado, A., ... & Mikosch, H. (2013). Integration of ligand and structure-based virtual screening for identification of leading anabolic steroids. *The Journal of Steroid Biochemistry and Molecular Biology*, *138*, 348-358. https://doi.org/10.1016/j.jsbmb.2013.07.004

48. Gayvert, K. M., Madhukar, N. S., & Elemento, O. (2016). A data-driven approach to predicting successes and failures of clinical trials. *Cell Chemical Biology*, *23*(10), 1294-1301. https://www.cell.com/cell-chemical-biology/pdf/S2451-9456(16)30291-4.pdf

49. Jorgensen, W. L., & Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences*, *102*(19), 6665-6670. https://doi.org/10.1073/pnas.0408037102

50. Sheng, C., Dong, G., Miao, Z., Zhang, W., & Wang, W. (2015). State-of-the-art strategies for targeting protein–protein interactions by small-molecule inhibitors. *Chemical Society Reviews*, *44*(22), 8238-8259. https://doi.org/10.1039/C5CS00252D

51. Wang, Y., Gao, J., Guo, X., Tong, T., Shi, X., Li, L., ... & Wang, H. (2014). Regulation of EGFR nanocluster formation by ionic protein-lipid interaction. *Cell Research*, *24*(8), 959-976. https://doi.org/10.1038/cr.2014.89

52. Alanis-Lobato, G., Andrade-Navarro, M. A., & Schaefer, M. H. (2016). HIPPIE v2. 0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic acids research*, gkw985. https://doi.org/10.1093/nar/gkw985

53. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: a free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, *52*(7), 1757-1768. https://doi.org/10.1021/ci3001277

54. Shen, M., Xiao, Y., Golbraikh, A., Gombar, V. K., & Tropsha, A. (2003). Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates. *Journal of Medicinal Chemistry*, *46*(14), 3013-3020. https://doi.org/10.1021/jm020491t

55. Unterthiner, T., Mayr, A., Klambauer, G., Steijaert, M., Wegner, J. K., Ceulemans, H., & Hochreiter, S. (2014, December). Deep learning as an opportunity in virtual screening. In *Proceedings of the deep learning workshop at NIPS* (Vol. 27, pp. 1-9). Cambridge, MA, United States: MIT Press. https://www.datascienceassn.org/sites/default/files/Deep%20Learning%20as%20an%20Opportunity%20in%20Virtual%20Screening.pdf

56. Polishchuk, P. (2017). Interpretation of quantitative structure–activity relationship models: past, present, and future. *Journal of Chemical Information and Modeling*, *57*(11), 2618-2639. https://doi.org/10.1021/acs.jcim.7b00274

57. Xu, Y., Ma, J., Liaw, A., Sheridan, R. P., & Svetnik, V. (2017). Demystifying multitask deep neural networks for quantitative structure–activity relationships. *Journal of Chemical Information and Modeling*, *57*(10), 2490-2504. https://doi.org/10.1021/acs.jcim.7b00087

58. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *64*, 4-17. https://doi.org/10.1016/j.addr.2012.09.019

59. Charifson, P. S., Corkery, J. J., Murcko, M. A., & Walters, W. P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry*, *42*(25), 5100-5109. https://doi.org/10.1021/jm990352k

60. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*, *3*(11), 935-949. https://doi.org/10.1038/nrd1549

61. Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews*, *66*(1), 334-395. https://doi.org/10.1124/pr.112.007336

62. Cross, S., Baroni, M., Goracci, L., & Cruciani, G. (2012). GRID-based three-dimensional pharmacophores I: FLAPpharm, a novel approach for pharmacophore elucidation. *Journal of Chemical Information and Modeling*, *52*(10), 2587-2598. https://doi.org/10.1021/ci300153d