



AI-Based Drug Discovery of TKIs Targeting L858R/T790M/C797S-Mutant EGFR in Non-small Cell Lung Cancer

Geunho Choi^{†*}, Daegeun Kim[†] and Junehwan Oh

AI LAB, AllLive Healthcare Co.,Ltd., Seongnam, Korea

OPEN ACCESS

Edited by:

Hideharu Kimura,
Kanazawa University, Japan

Reviewed by:

Muthupandian Saravanan,
Saveetha Dental College And
Hospitals, India
Prashant S Kharkar,
Institute of Chemical Technology, India

*Correspondence:

Geunho Choi
cgh2797@naver.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Pharmacology of Anti-Cancer Drugs,
a section of the journal
Frontiers in Pharmacology

Received: 29 January 2021

Accepted: 19 July 2021

Published: 28 July 2021

Citation:

Choi G, Kim D and Oh J (2021) AI-
Based Drug Discovery of TKIs
Targeting L858R/T790M/C797S-
Mutant EGFR in Non-small Cell
Lung Cancer.
Front. Pharmacol. 12:660313.
doi: 10.3389/fphar.2021.660313

Lung cancer has a high mortality rate, and non-small cell lung cancer (NSCLC) is the most common type of lung cancer. Patients have been observed to acquire resistance against various anticancer agents used for NSCLC due to L858R (or Exon del19)/T790M/C797S-EGFR mutations. Therefore, next-generation drugs are being developed to overcome this problem of acquired resistance. The goal of this study was to use artificial intelligence (AI) to discover drug candidates that can overcome acquired resistance and reduce the limitations of the current drug discovery process, such as high costs and long durations of drug design and production. To generate ligands using AI, we collected data related to tyrosine kinase inhibitors (TKIs) from accessible libraries and used LSTM (Long short term memory) based transfer learning (TL) model. Through the simplified molecular-input line-entry system (SMILES) datasets of the generated ligands, we obtained drug-like ligands via parameter-filtering, cyclic skeleton (CSK) analysis, and virtual screening utilizing deep-learning method. Based on the results of this study, we are developing prospective EGFR TKIs for NSCLC that have overcome the limitations of existing third-generation drugs.

Keywords: NSCLC, EGFR, tyrosine kinase inhibitors (TKIs), transfer learning, LSTM, virtual screening

INTRODUCTION

Cancer is one of the leading causes of death worldwide (Sung et al., 2021). In the United States, cancer is the second leading cause of death as of 2017–2018, accounting for 21% of the deaths. The mortality rates according to the cancer type from 2014 to 2018 in the United States are as follows (rates are per 100,000 population). Lung and bronchus cancer was 38.5, breast cancer (female) was 20.1, prostate cancer was 19.0, colon and rectum cancer was 13.7, liver and intrahepatic bile duct cancer was 6.6, and Kidney and renal pelvis cancer was 3.6 (Siegel et al., 2021). Among many cancers, lung cancer has a high mortality rate. Approximately 1.7 million people died from lung cancer in 2018, with non-small cell lung cancer (NSCLC) being the cause of death in over 80% of these cases (Yuan et al., 2019). Platinum-based doublet chemotherapy (e.g., cisplatin), which directly induces cancer cell apoptosis, has been used to treat lung cancer since 2005. However, this treatment has drawbacks such as inducement of normal cell apoptosis and achievement of only a short survival period of 10 months. Since the discovery of genetic mutations related to epidermal growth factor receptor (EGFR) in 2004, numerous targeted anticancer agents with fewer side effects compared to chemotherapy have been developed. These agents have extended the survival period of patients with lung cancer to over 24 months (Li et al., 2019) (Jiao et al., 2018) (Yuan et al., 2019).

Ongoing research on targeted anticancer agents has revealed the mechanisms of various targetable pathways associated with lung cancer (e.g., EGFR, PI3K/AKT/mTOR, RAS-MAPK, and JAK/STAT). EGFR, a member of the HER family, is a transmembrane glycoprotein that regulates cell regulatory pathways involved in cell proliferation, differentiation, and apoptosis. Since the discovery of EGFR overexpression in patients with lung cancer, which revealed a correlation between EGFR tyrosine kinase expression and tumor formation, numerous agents with significant therapeutic targets in NSCLC have been developed (Le and Gerber, 2019) (Leonetti et al., 2019).

Gefitinib from AstraZeneca and Erlotinib from Roche, which are first-generation reversible inhibitors of EGFR tyrosine kinase, were approved in 2003–2013, respectively. Afatinib, a second-generation irreversible inhibitor of EGFR tyrosine kinase from Boehringer Ingelheim, was approved in 2013. These first- and second-generation drugs are used as targeted agents for NSCLC and have shown high efficacy for common activating EGFR mutations such as the L858R point mutation and exon 19 deletion. However, after a treatment period of 1–2 years, the second mutation called the “gatekeeper”, referring to the T790M-mutation in EGFR exon 20, occurs in 50–60% of patients treated with these agents, in addition to other mutations such as MET amplification and RAS mutations. The “gatekeeper” mutation reduces the effectiveness of the first- and second-generation anticancer agents by inducing drug resistance. Thus, numerous third-generation EGFR tyrosine kinase inhibitors (TKIs) sensitive to TK domain mutations (T790M) have been developed. (Jett and Carr, 2013) (Yuan et al., 2019) (Liu et al., 2018) (Grabe et al., 2018).

Osimertinib (Tagrisso) is a major third-generation EGFR TKI developed by AstraZeneca and approved by the Food and Drug Administration (FDA) in 2015. Osimertinib covalently binds (Ghosh et al., 2019) to the Cys797 residue of EGFR tyrosine kinase and is thus highly selective (Zhai et al., 2020) (Klaeger et al., 2017) and potent for the EGFR T790M mutation and other activating EGFR mutations. In 2015, however, the use of osimertinib as a third-generation EGFR TKI was shown to lead to acquired resistance, resulting from the tertiary point mutation C797S. Substitution of the Cys797 residue with serine 797 led to the loss of covalent interactions and significantly reduced drug efficacy. Consequently, fourth-generation drugs with therapeutic effects against the EGFR C797S-mutation are currently under development. (Jett and Carr, 2013) (Leonetti et al., 2019) (Grabe et al., 2018).

Drug discovery costs are increasing and research and development efficiency is decreasing (Mak and Pichika, 2019) (Scannell et al., 2012) (Schuhmacher et al., 2016). Therefore, increasing efforts have been undertaken to use artificial intelligence (AI) in drug discovery (Chen et al., 2018) (Chan et al., 2019). Unlike conventional drug discovery procedures, AI based drug discovery does not incur high experimental costs and requires only a small number of personnel.

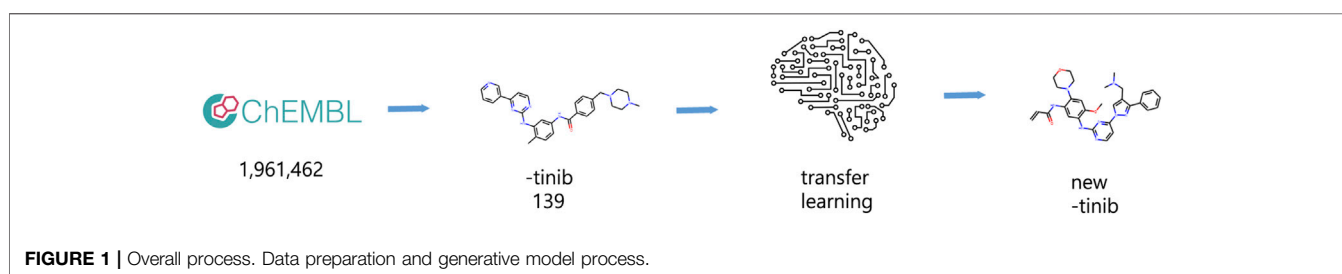
Deep learning (Lecun et al., 2015) is artificial neural networks that mimic the brain, a complex system. Deep learning has been successfully applied to areas such as computer vision (Voulodimos et al., 2018), speech recognition (Nassif et al., 2019), and natural language processing (Young et al., 2018). Recently, studies applying AI such as deep learning to drug discovery are increasing. Researchers have developed a drug generation model using variational autoencoder (Gómez-Bombarelli et al., 2018), generative adversarial autoencoder models (Kadurin et al., 2017). A drug generation model using a recurrent neural network (RNN) architecture and reinforcement learning (Popova et al., 2018) has also been developed. Deep learning is highly sensitive to data quality and quantity. A small dataset is a bottleneck in AI-aided novel drug discovery and can be overcome by transfer learning (TL) (Segler et al., 2018) (Gupta et al., 2018) (Moret et al., 2020) (Cai et al., 2020). TL enables efficient learning even with a small amount of data.

We adopted model (Li et al., 2020) using RNN(LSTM) and TL, and conducted research with the aim of discovering 4th generation new drug candidates as L858R (or Exon del19)/T790M/C797S-mutation EGFR tyrosine kinase inhibitors related to NSCLC.

MATERIALS AND METHODS

Data Curation and Analysis

We downloaded data for 1,961,462 compounds from ChEMBL (Gaulton et al., 2012), a curated compound database, and selected compounds whose names ended with ‘-tinib’ and additionally selected Lazertinib, creating a list of 139 compounds (Figure 1). The reason why we specifically chose ‘-tinib’ structures as our base dataset molecules from ChEMBL database is that ‘-tinib’ is an already known tyrosine kinase inhibitor that exhibits certain pharmacological effects in relation to various tyrosine kinases including our targeted protein, EGFR TK and We aim to discover promising



candidates as EGFR TKIs reflecting the structural, physicochemical and biochemical features of these ‘-tinib’. The data used in the paper are available at https://github.com/cgh2797/AI_drug_discovery_EGFR. Data was input in SMILES format using the open-source cheminformatics Rdkit 2020.03.1.0. We performed a 10-fold augmentation on the ‘-tinib’ dataset as it was not large enough to train a model (Bjerrum, 2017). Additionally, we analyzed the structural similarity between compounds by examining their cyclic skeletons (CSKs) (Xu and Johnson, 2002).

LSTM TL

Since the dataset is small and SMILES is a string format, an RNN (LSTM) TL model (Li et al., 2020) was selected. A training dataset (Li et al., 2020) was used as a base dataset, and the dataset of 139 ‘-tinib’ compounds was used as a second dataset for TL after 10-fold augmentation. The data preprocessing method was selected from the previous study (Li et al., 2020). The BasicLSTMCell function in TensorFlow was used for the two LSTM layers of the deep learning model. A dropout was applied to each LSTM layer. The keep probability was 0.8, and the number of hidden layers was 512. For the loss function, TensorFlow’s seq2seq.sequence_loss function optimized with the Adam optimizer was used. The learning rate was set to

0.003. Model training was performed using TensorFlow-gpu 1.15.0. NVIDIA GeForce RTX 2080 SUPER was used for computation.

Filtering

Of the generated molecules, invalid molecules whose parameters could not be calculated by Rdkit were filtered out. Next, parameters including molecular weight, LogP, HBA, HBD, TPSA, and rotatable bonds were calculated using Rdkit. The weighted mean of the quantitative estimates of drug-likeness (QED) (Bickerton et al., 2012) was calculated using Rdkit. The desirability functions (d) can be described as asymmetric double sigmoidal (ADS) functions and are expressed as shown in Eq. 1. a, b, c, d, e, and f in Eq. 1 denote the parameters of the ADS function. QED is calculated by taking the geometric mean of the desirability functions multiplied by their weights w and can be expressed as shown in Eq. 2. Expanding the equation results in Eq. 3: This material is from our original study (Bickerton et al., 2012).

$$D(x) = a + \frac{b}{\left[1 + \exp\left(-\frac{x-c+\frac{d}{2}}{e}\right)\right]} \left[1 - \frac{1}{\left[1 + \exp\left(-\frac{x-c-\frac{d}{2}}{f}\right)\right]}\right] \quad (1)$$

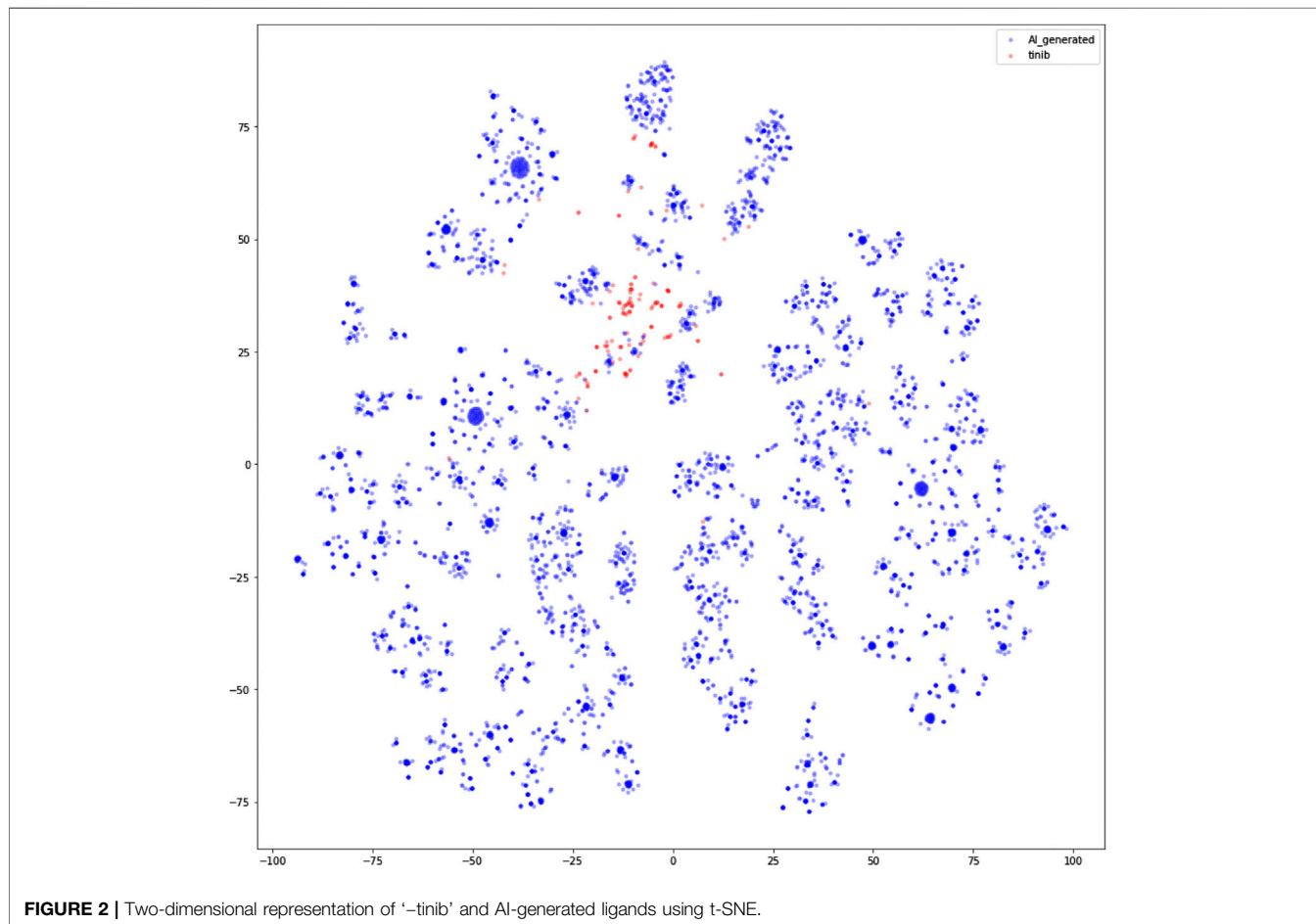


FIGURE 2 | Two-dimensional representation of ‘-tinib’ and AI-generated ligands using t-SNE.

$$QED_W = \exp\left(\frac{\sum_{i=1}^n w_i \ln d_i}{\sum_{i=1}^n w_i}\right) \quad (2)$$

$$QED_W = \exp\left[\frac{W_{MW} \ln d_{MW} + W_{ALOGP} \ln d_{ALOGP} + W_{HBA} \ln d_{HBA} + W_{HBD} \ln d_{HBD} + W_{PSA} \ln d_{PSA} + W_{ROTB} \ln d_{ROTB} + W_{AROM} \ln d_{AROM} + W_{ALERTS} \ln d_{ALERTS}}{W_{MW} + W_{ALOGP} + W_{HBA} + W_{HBD} + W_{PSA} + W_{ROTB} + W_{AROM} + W_{ALERTS}}\right] \quad (3)$$

The synthetic accessibility (SA) (Ertl and Schuffenhauer, 2009) score was calculated as a combination of two components using the Rdkit.

$$Sascore = \text{fragmentScore} - \text{complexityPenalty} \quad (4)$$

The following screening filters were used for the parameters: $300 \leq MW \leq 700$, $2.0 \leq \text{LogP} \leq 6.0$, $2.0 \leq \text{HBD} \leq 6.0$, $0 \leq \text{HBA} \leq 12.0$, $\text{HBA} + \text{HBD} \leq 14.0$, $60.0 \leq \text{TPSA} \leq 140.0$, and rotational bond ≤ 12.0 . All filters were applied to obtain the desirable ligands.

AI Virtual Screening

In virtual screening, DeepDTA (Öztürk et al., 2018), a convolutional neural network-based drug target affinity prediction model, was used to predict the affinity of the candidate compounds for L858R/T790M/C797S mutant EGFR (PDB code: 6LUD). The output of the model is pK_d (5), which denotes the affinity between a protein and a drug.

$$\text{PK}_d = -\log_{10}\left(\frac{K_d}{1e^9}\right) \quad (5)$$

pK_d was predicted using Tensorflow 2.2.0, keras 2.4.3, and NVIDIA GeForce RTX 2080 SUPER.

RESULTS

AI-Aided Drug Discovery

TL was used to compensate for the small quantity of ‘-tinib’ data obtained from ChEMBL. The compounds generated via TL had unique structures that were similar to the ‘-tinib’, but also exhibited the characteristics of the compounds in the training dataset. The compounds were vectorized using the Morgan Fingerprint (Cereto-Massagué et al., 2015) in Rdkit and visualized after dimensionality reduction into a two-dimensional (2D) space using t-SNE (Maaten and Hinton, 2008) in scikit-learn (Figure 2). In the 2D space, compounds with similar structures were clustered closer together, while structurally dissimilar compounds were farther from one another. The AI-generated compounds surrounded the ‘-tinib’ compounds in a ring shape. While the AI-generated compounds were similar to the ‘-tinib’ compounds based on their small distance between one another in the 2D space, they were still far enough to be considered unique, and thereby avoided patent infringement.

Parameter Filtering

Approximately 20% of 10,316 AI-generated ligand SMILES were invalid SMILES that did not meet the encoding rules and were thus removed. Following the removal, the remaining ligands were screened by filtering based on MW, LogP, TPSA, HBA, HBD,

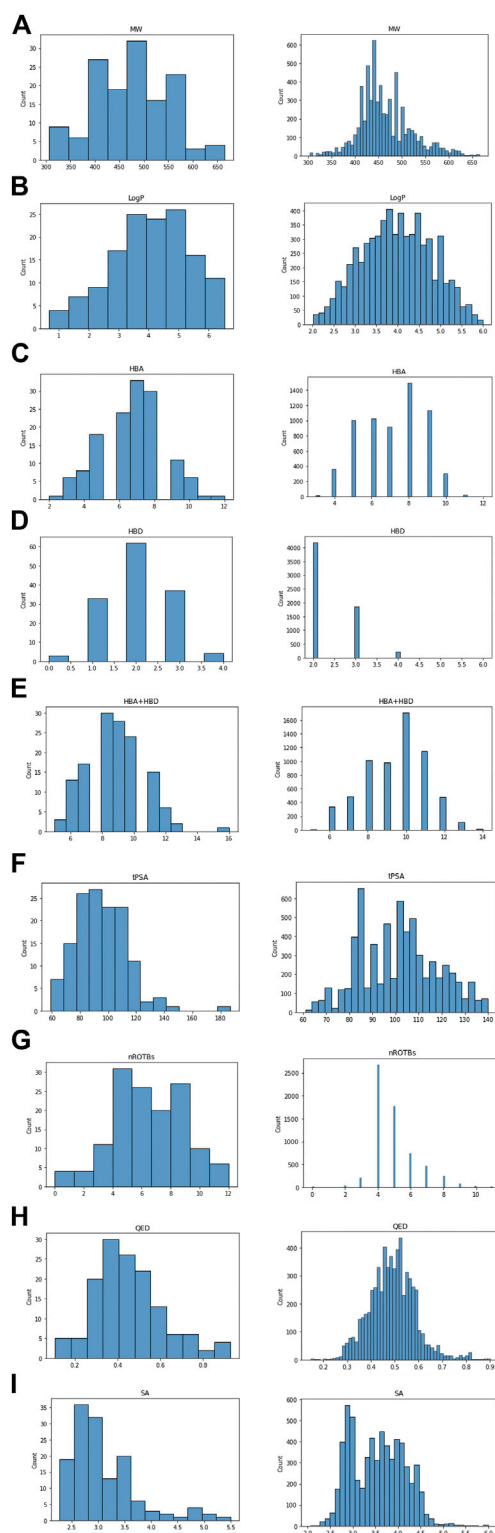
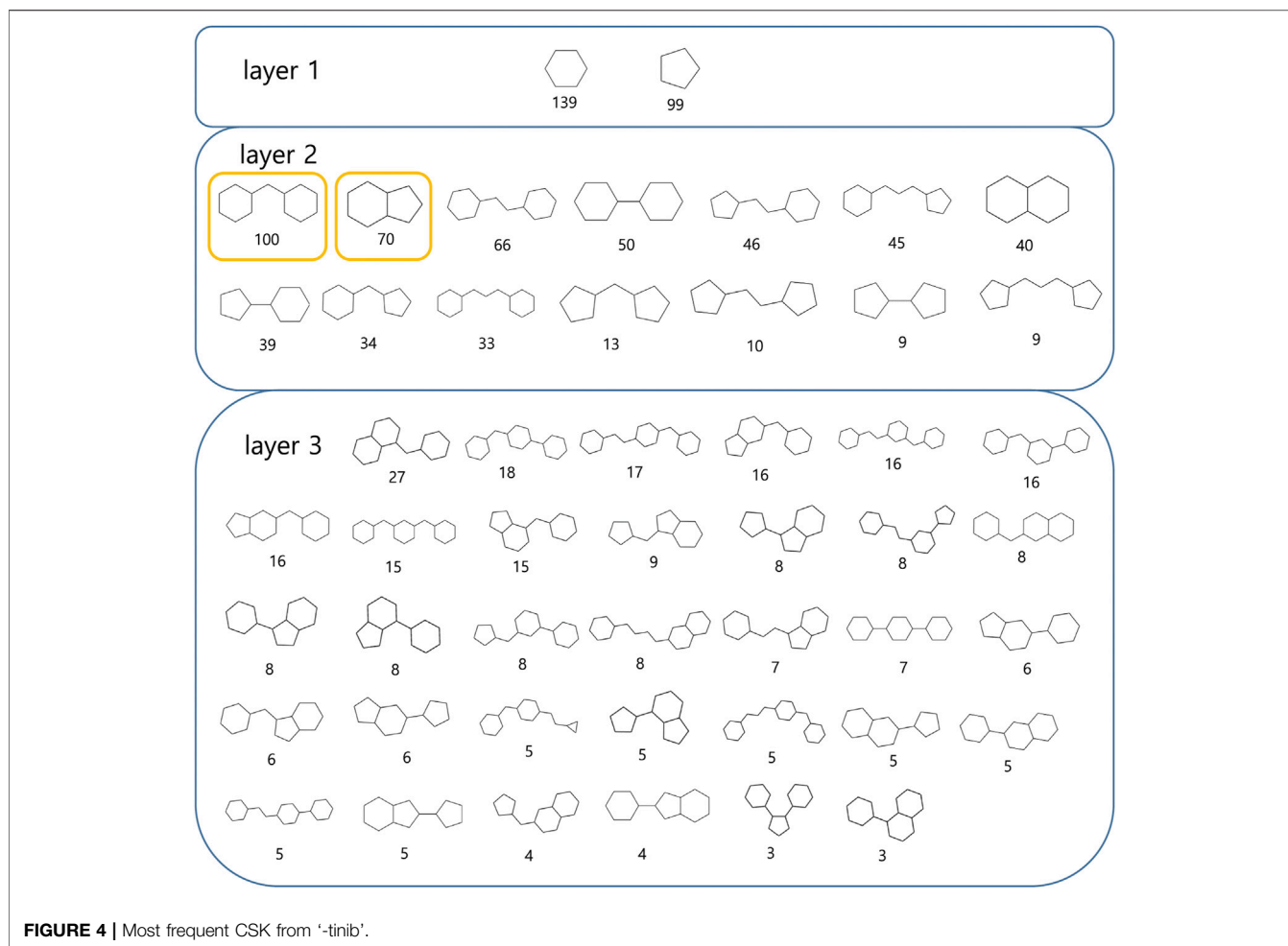


FIGURE 3 | Comparison parameter distribution of ‘-tinib’ and 6,283 ligands made by AI. The figure shows the parameter distribution, the ‘-tinib’ on the left and the AI generated ligands on the right. **(A)** Molecular Weight, **(B)** LogP, **(C)** HBA, **(D)** HBD, **(E)** HBA + HBD, **(F)** tPSA, **(G)** Rotatable bond, **(H)** QED, and **(I)** SA (Synthetic Accessibility) score.



HBA + HBD, and rotatable bonds to ultimately remove undruggable molecules. As a result of screening using parameter filtering, we obtain 6,283 ligands out of 10,316. After that, the distributions of parameters such as MW, LogP, TPSA, HBA, HBD, HBA + HBD, rotatable bond, QED, and SA were compared between the AI-generated and '-tinib' compounds (Figure 3). The parameters showed highly similar distributions between the two groups of compounds. The physicochemical characteristics of the existing drugs were well-reproduced by the AI-generated compounds.

Structural Similarity Based on CSK

A scaffold is the core structure of a compound. CSK is an abstract version of a scaffold. We examined pharmaceutically meaningful structural similarities between '-tinib' by computing CSKs to select drug-like compounds. We created a hierarchical figure by placing structures with a single ring in layer 1, those with two rings in layer 2, and those with three rings in layer 3 (Figure 4).

Bridged-bicyclic rings and fused-bicyclic rings were the two most commonly observed types of CSK, each, with a count of 100 and 70 (marked with yellow). To assess the reproducibility of

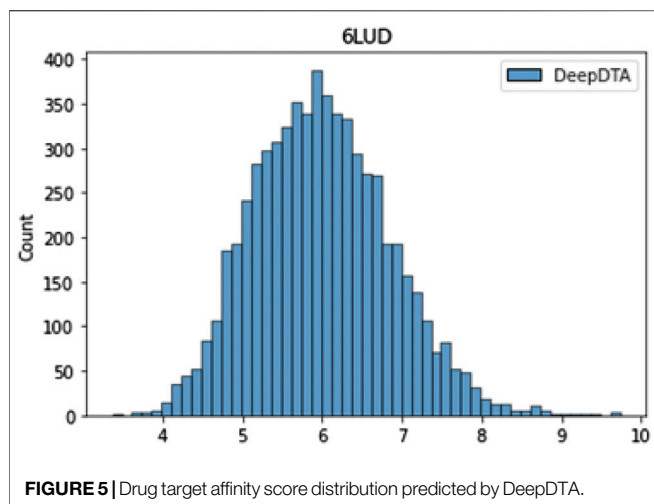
these results, we analyzed the CSKs of 6,283 molecules generated from the training dataset of 139 '-tinib' compounds using a machine learning model. The 3,308 and 2,254 ligands had bridged-bicyclic rings and fused-bicyclic rings, respectively, and thus, the results were deemed reproducible. We confirmed the structural similarities between the original and AI-generated ligand groups based on CSKs.

AI Virtual Screening

We used DeepDTA, a machine learning-based model, for fast virtual screening of druggable ligands based on their target binding affinity for L858R/T790M/C797S mutant EGFR (PDB code: 6LUD). Figure 5 shows the affinity distribution predicted by DeepDTA. Ligands with high pK_d score were predicted to have high affinity for the L858R/T790M/C797S mutant EGFR.

Screening Ligands Based on Stringent Criterion

To extract ligands for docking simulation and non-clinical experiments, we screened ligands using stringent criterion. As a result, 360 ligands that we can calculate by docking program



were selected. More information on 360 ligands is available at https://github.com/cgh2797/AI_drug_discovery_EGFR.

DISCUSSION

In this study, we used AI-based drug discovery to overcome the issues of high cost and low efficiency of drug research and development at present. We used AI to discover drug-like ligands resembling TKIs associated with EGFR in NSCLC and screened the candidates through the following process.

We extracted 139 ligands associated with TKIs from approximately 1.96 million compounds in ChEMBL. Next, we performed deep learning using an RNN (LSTM) to generate 10,316 SMILES associated with TKI molecules. Through parameter-filtering using in-house methods, we narrowed down the SMILES to 6,283 drug-like ligands with affinity for L858R/T790M/C797S mutant EGFR in NSCLC. To gain additional understanding of the selected ligands, we analyzed their CSKs to examine the structural similarity between the AI-generated molecules and the existing ‘-tinib’ from ChEMBL. We used a deep learning model such as DeepDTA to predict the binding affinities of these compounds for L858R/T790M/C797S mutant EGFR. Finally, by applying stringent criterion, 360 ligands were obtained.

However, there are several limitations to this study. First, it is difficult to create only covalent ligands or determine if the ligands are covalent, when generating various ligands through AI methods. Since we based our results on Osimertinib, which is a representative covalent TKI, obtaining covalently binding ligands is also an ideal aim in this study. Thus, in a follow-up

REFERENCES

- Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., and Hopkins, A. L. (2012). Quantifying the Chemical beauty of Drugs. *Nat. Chem.* 4, 90–98. doi:10.1038/nchem.1243
- Bjerrum, E. J. (2017). *Smiles Enumeration as Data Augmentation for Neural Network Modeling of Molecules*. arXiv.

study, we will select promising compounds by determining directly based on scientific rationale whether the 360 ligands are “covalent” or “noncovalent” ligands using an in silico docking prediction method. Second, there are selectivity issues, which should be addressed even within the tyrosine kinase family. Since our research is in its early stage, our primary goal of the research is to preferentially discovery candidates as EGFR TKIs with notable efficacy (i.e., binding affinity). After identifying promising candidates, a more detailed research would be conducted to resolve the selectivity issues. Finally, another limitation of this study is that although new druggable ligands were found, experiments such as in silico docking, synthesis in the laboratory, and preclinical trials were not conducted. Hence, further studies on improving the therapeutic potential of our selected ligands, such as in silico docking prediction, synthesis in the laboratory, and preclinical trials (e.g., efficacy and safety trials), would be undertaken. Accordingly, a more detailed research would also be conducted to resolve the aforementioned tyrosine kinase selectivity issues (e.g., SAR by structural modification).

We don’t put meaning to simply discover new druggable ligands similar to existing ‘-tinib’ using AI. Since AI has infinite potential for applications in drug discovery, our goal is not only limited to drug discovery, but also includes the successful development of drugs that can receive FDA approval.

Therefore, our next task is to discover new candidates with good drug-like profiles (efficacy, toxicity, pharmacodynamics, pharmacokinetics, etc.) and identify those eligible for drug approval. We must also explore the possibility of using these compounds in combination therapy. By presenting examples of successful new drug development through these series of processes, new drug development technology using AI will become a new drug discovery paradigm.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GC and JO designed the study. All experiments were performed by GC. Pharmaceutical analyzes were performed by DK. GC and DK contributed to the writing of the manuscript.

- Cai, C., Wang, S., Xu, Y., Zhang, W., Tang, K., Ouyang, Q., et al. (2020). Transfer Learning for Drug Discovery. *J. Med. Chem.* 63, 8683–8694. doi:10.1021/acs.jmedchem.9b02147
- Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., and Pujadas, G. (2015). Molecular Fingerprint Similarity Search in Virtual Screening. *Methods* 71, 58–63. doi:10.1016/j.ymeth.2014.08.005
- Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., and Yuan, S. (2019). Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* 40, 592–604. doi:10.1016/j.tips.2019.06.004

- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018). The Rise of Deep Learning in Drug Discovery. *Drug Discov. Today* 23, 1241–1250. doi:10.1016/j.drudis.2018.01.039
- Ertl, P., and Schuffenhauer, A. (2009). Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminform.* 1, 1–11. doi:10.1186/1758-2946-1-8
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* 40, D1100–D1107. doi:10.1093/nar/gkr777
- Ghosh, A. K., Samanta, I., Mondal, A., and Liu, W. R. (2019). Covalent Inhibition in Drug Discovery. *ChemMedChem* 14, 889–906. doi:10.1002/cmdc.201900107
- Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., et al. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* 4, 268–276. doi:10.1021/acscentsci.7b00572
- Grabe, T., Lategahn, J., and Rauh, D. (2018). C797S Resistance: The Undruggable EGFR Mutation in Non-small Cell Lung Cancer?. *ACS Med. Chem. Lett.* 9, 779–782. doi:10.1021/acsmchemlett.8b00314
- Gupta, A., Müller, A. T., Huisman, B. J. H., Fuchs, J. A., Schneider, P., and Schneider, G. (2018). Generative Recurrent Networks for De Novo Drug Design. *Mol. Inf.* 37, 1700111. doi:10.1002/minf.201700111
- Jett, J. R., and Carr, L. L. (2013). Targeted Therapy for Non-small Cell Lung Cancer. *Am. J. Respir. Crit. Care Med.* 188, 907–912. doi:10.1164/rccm.201301-0189PP
- Jiao, Q., Bi, L., Ren, Y., Song, S., Wang, Q., and Wangshan, Y.-s. (2018). Advances in Studies of Tyrosine Kinase Inhibitors and Their Acquired Resistance. *Mol. Cancer* 17, 1–12. doi:10.1186/s12943-018-0801-5
- Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., and Zhavoronkov, A. (2017). DruGAN: An Advanced Generative Adversarial Autoencoder Model for De Novo Generation of New Molecules with Desired Molecular Properties In Silico. *Mol. Pharmaceutics* 14, 3098–3104. doi:10.1021/acsmolpharmaceut.7b00346
- Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P.-A., et al. (2017). The Target Landscape of Clinical Kinase Drugs. *Science* 358, eaan4368. doi:10.1126/science.aan4368
- Le, T., and Gerber, D. (2019). Newer-generation EGFR Inhibitors in Lung Cancer: How Are They Best Used? *Cancers* 11(3), 366. doi:10.3390/cancers11030366
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521, 436–444. doi:10.1038/nature14539
- Leonetti, A., Sharma, S., Minari, R., Perego, P., Giovannetti, E., and Tiseo, M. (2019). Resistance Mechanisms to Osimertinib in EGFR-Mutated Non-small Cell Lung Cancer. *Br. J. Cancer* 121, 725–737. doi:10.1038/s41416-019-0573-8
- Li, Q., Zhang, T., Li, S., Tong, L., Li, J., Su, Z., et al. (2019). Discovery of Potent and Noncovalent Reversible EGFR Kinase Inhibitors of EGFR L858R/T790M/C797S. *ACS Med. Chem. Lett.* 10, 869–873. doi:10.1021/acsmchemlett.8b00564
- Li, X., Xu, Y., Yao, H., and Lin, K. (2020). Chemical Space Exploration Based on Recurrent Neural Networks: Applications in Discovering Kinase Inhibitors. *J. Cheminform.* 12, 1–13. doi:10.1186/s13321-020-00446-3
- Liu, Q., Yu, S., Zhao, W., Qin, S., Chu, Q., and Wu, K. (2018). EGFR-TKIs Resistance via EGFR-independent Signaling Pathways. *Mol. Cancer* 17, 1–9. doi:10.1186/s12943-018-0793-1
- Maaten, L., and Hinton, G. (2008). Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Mak, K.-K., and Pichika, M. R. (2019). Artificial Intelligence in Drug Development: Present Status and Future Prospects. *Drug Discov. Today* 24, 773–780. doi:10.1016/j.drudis.2018.11.014
- Moret, M., Friedrich, L., Grisoni, F., Merk, D., and Schneider, G. (2020). Generative Molecular Design in Low Data Regimes. *Nat. Mach. Intell.* 2, 171–180. doi:10.1038/s42256-020-0160-y
- Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., and Shaalan, K. (2019). Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access* 7, 19143–19165. doi:10.1109/ACCESS.2019.2896880
- Öztürk, H., Özgür, A., and Ozkirimli, E. (2018). DeepDTA: Deep Drug-Target Binding Affinity Prediction. *Bioinformatics* 34, i821–i829. doi:10.1093/bioinformatics/bty593
- Popova, M., Isayev, O., and Tropsha, A. (2018). Deep Reinforcement Learning for De Novo Drug Design. *Sci. Adv.* 4, eaap7885–15. doi:10.1126/sciadv.aap7885
- Scannell, J. W., Blanckley, A., Boldon, H., and Warrington, B. (2012). Diagnosing the Decline in Pharmaceutical R&D Efficiency. *Nat. Rev. Drug Discov.* 11, 191–200. doi:10.1038/nrd3681
- Schuhmacher, A., Gassmann, O., and Hinder, M. (2016). Changing R&D Models in Research-Based Pharmaceutical Companies. *J. Transl. Med.* 14, 1–11. doi:10.1186/s12967-016-0838-4
- Segler, M. H. S., Kogej, T., Tyrchan, C., and Waller, M. P. (2018). Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* 4, 120–131. doi:10.1021/acscentsci.7b00512
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660
- Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Comput. Intelligence Neurosci.* 2018, 1–13. doi:10.1155/2018/7068349
- Xu, Y.-J., and Johnson, M. (2002). Using Molecular Equivalence Numbers to Visually Explore Structural Features that Distinguish Chemical Libraries. *J. Chem. Inf. Comput. Sci.* 42, 912–926. doi:10.1021/ci025535l
- Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. *IEEE Comput. Intell. Mag.* 13, 55–75. doi:10.1109/MCI.2018.2840738
- Yuan, M., Huang, L.-L., Chen, J.-H., Wu, J., and Xu, Q. (2019). The Emerging Treatment Landscape of Targeted Therapy in Non-small-cell Lung Cancer. *Sig Transduct Target. Ther.* 4. doi:10.1038/s41392-019-0099-9
- Zhai, X., Ward, R. A., Doig, P., and Argyrou, A. (2020). Insight into the Therapeutic Selectivity of the Irreversible EGFR Tyrosine Kinase Inhibitor Osimertinib through Enzyme Kinetic Studies. *Biochemistry* 59, 1428–1441. doi:10.1021/acsbiochem.0c00104

Conflict of Interest: Authors GC, DK and JO were employed by AllLive Healthcare Co., Ltd.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Choi, Kim and Oh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.