References

1. Eder, J. and P.L. Herrling, *Trends in Modern Drug Discovery.* Handb Exp Pharmacol, 2016. 232: p. 3-22.
2. Sacan, A., S. Ekins, and S. Kortagere, *Applications and limitations of in silico models in drug discovery.* Methods Mol Biol, 2012. 910: p. 87-124.
3. Shakil, S. and M.F. Abuzinadah, *Putative Anti-Cancer Drug Candidate Targeting the 'PLK-1-Polo-Box Domain' by High Throughput Virtual Screening: A Computational Drug Design Study.* Crit Rev Eukaryot Gene Expr, 2019. 29(3): p. 251-261.
4. Ashtekar, S.S., N.M. Bhatia, and M.S. Bhatia, *Development of leads targeting ER-alpha in breast cancer: An in silico exploration from natural domain.* Steroids, 2018. 131: p. 14-22.
5. McCulloch, W.S. and W. Pitts, *A logical calculus of the ideas immanent in nervous activity. 1943.* Bull Math Biol, 1990. 52(1-2): p. 99-115; discussion 73-97.
6. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. 521(7553): p. 436-44.
7. Winkler, D.A. and T.C. Le, *Performance of Deep and Shallow Neural Networks, the Universal Approximation Theorem, Activity Cliffs, and QSAR.* Mol Inform, 2017. 36(1-2).
8. Wang, Y. and J. Zeng, *Predicting drug-target interactions using restricted Boltzmann machines.* Bioinformatics, 2013. 29(13): p. i126-34.
9. Ma, J., et al., *Deep neural nets as a method for quantitative structure-activity relationships.* J Chem Inf Model, 2015. 55(2): p. 263-74.
10. Schmidhuber, J., *Deep learning in neural networks: an overview.* Neural Netw, 2015. 61: p. 85-117.
11. Nosengo, N., *Can you teach old drugs new tricks?* Nature, 2016. 534(7607): p. 314-6.

12. Wouters, O.J., M. McKee, and J. Luyten, *Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018.* JAMA, 2020. 323(9): p. 844-853.

13. Darrow, J.J., J. Avorn, and A.S. Kesselheim, *FDA Approval and Regulation of Pharmaceuticals, 1983-2018.* JAMA, 2020. 323(2): p. 164-176.

14. Dickson, M. and J.P. Gagnon, *Key factors in the rising cost of new drug discovery and development.* Nat Rev Drug Discov, 2004. 3(5): p. 417-29.

15. Lavecchia, A. and C. Di Giovanni, *Virtual screening strategies in drug discovery: a critical review.* Curr Med Chem, 2013. 20(23): p. 2839-60.

16. Chaguturu, R., *Combinatorial Chemistry & High Throughput Screening. Editorial.* Comb Chem High Throughput Screen, 2013. 16(1): p. 1.

17. Lavecchia, A., *Machine-learning approaches in drug discovery: methods and applications.* Drug Discov Today, 2015. 20(3): p. 318-31.

18. Devi, R.V., S.S. Sathya, and S.M. Coumar, *Multi-objective Genetic Algorithm for De Novo Drug Design.* Curr Comput Aided Drug Des, 2020.

19. Popova, M., O. Isayev, and A. Tropsha, *Deep reinforcement learning for de novo drug design.* Sci Adv, 2018. 4(7): p. eaap7885.

20. Gomez-Bombarelli, R., et al., *Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules.* ACS Cent Sci, 2018. 4(2): p. 268-276.

21. Kang, S. and K. Cho, *Conditional Molecular Design with Deep Generative Models.* J Chem Inf Model, 2019. 59(1): p. 43-52.

22. Li, Y., L. Zhang, and Z. Liu, *Multi-objective de novo drug design with conditional graph generative model.* J Cheminform, 2018. 10(1): p. 33.

23. Stahl, N., et al., *Deep Reinforcement Learning for Multiparameter Optimization in de novo Drug Design.* J Chem Inf Model, 2019. 59(7): p. 3166-3176.

24. Charest, A., et al., *ROS fusion tyrosine kinase activates a SH2 domain-containing phosphatase-2/phosphatidylinositol 3-kinase/mammalian target of rapamycin*

signaling axis to form glioblastoma in mice. Cancer Res, 2006. 66(15): p. 7473-81.

25. Patil, T., et al., *Targeted therapies for ROS1-rearranged non-small cell lung cancer.* Drugs Today (Barc), 2019. 55(10): p. 641-652.
26. Shaw, A.T., et al., *Crizotinib in ROS1-rearranged non-small-cell lung cancer.* N Engl J Med, 2014. 371(21): p. 1963-71.
27. Gainor, J.F., et al., *Patterns of Metastatic Spread and Mechanisms of Resistance to Crizotinib in ROS1-Positive Non-Small-Cell Lung Cancer.* JCO Precis Oncol, 2017. 2017.
28. Gainor, J.F. and A.T. Shaw, *Novel targets in non-small cell lung cancer: ROS1 and RET fusions.* Oncologist, 2013. 18(7): p. 865-75.
29. Uguen, A. and M. De Braekeleer, *ROS1 fusions in cancer: a review.* Future Oncol, 2016. 12(16): p. 1911-28.
30. Seeliger, D. and B.L. de Groot, *Ligand docking and binding site analysis with PyMOL and Autodock/Vina.* J Comput Aided Mol Des, 2010. 24(5): p. 417-22.
31. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. 25(13): p. 1605-12.
32. Koes, D.R. and C.J. Camacho, *ZINCPharmer: pharmacophore search of the ZINC database.* Nucleic Acids Res, 2012. 40(Web Server issue): p. W409-14.
33. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. 31(2): p. 455-61.
34. Daina, A., O. Michielin, and V. Zoete, *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.* Sci Rep, 2017. 7: p. 42717.
35. Sanner, M.F., A.J. Olson, and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces.* Biopolymers, 1996. 38(3): p. 305-20.

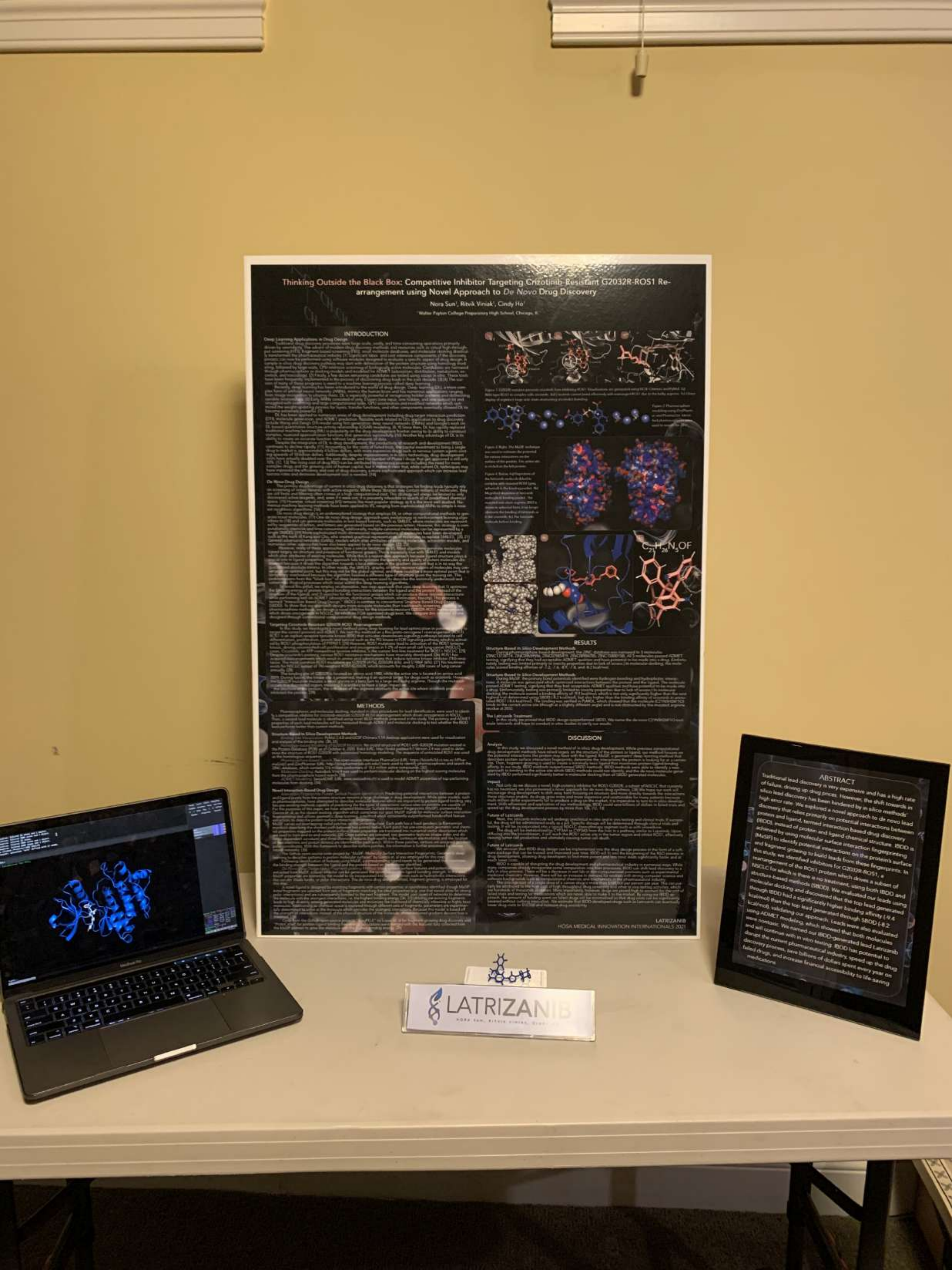
36. Gainza, P., et al., *Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning.* Nat Methods, 2020. **17**(2): p. 184-192.

37. Wang, L., J. Chambers, and R. Abel, *Protein-Ligand Binding Free Energy Calculations with FEP.* Methods Mol Biol, 2019. **2022**: p. 201-232.

38. Jespers, W., J. Aqvist, and H. Gutierrez-de-Teran, *Free Energy Calculations for Protein-Ligand Binding Prediction.* Methods Mol Biol, 2021. **2266**: p. 203-226.

39. Fischer, T., S. Gazzola, and R. Riedl, *Approaching Target Selectivity by De Novo Drug Design.* Expert Opin Drug Discov, 2019. **14**(8): p. 791-803.

40. Perez, C., et al., *FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design.* J Chem Inf Model, 2020. **60**(3): p. 1728-1736.


Thinking Outside the Black Box: Competitive Inhibitor Targeting Crizotinib-Resistant G2032R-ROS1 Re-arrangement using Novel Approach to De Novo Drug Discovery
Nora Sun[1], Ritvik Viniak[1], Cindy Ho[1]
[1]Walter Payton College Preparatory High School, Chicago, IL
INTRODUCTION
METHODS
RESULTS
DISCUSSION
LATRIZANIB
HOSA MEDICAL INNOVATION INTERNATIONALS 2021
ABSTRACT
Traditional lead discovery is very expensive and has a high rate of failure, driving up drug prices. However, the shift towards in silico lead discovery has been hindered by in silico methods' high error rate. We explored a novel approach to de novo lead discovery that relies primarily on potential interactions between protein and ligand, termed interaction based drug discovery (IBDD), instead of protein and ligand chemical structure. IBDD is achieved by using molecular surface interaction fingerprinting (MaSIF) to identify potential interactions on the protein's surface and fragment growing to build leads from these fingerprints. In this study, we identified inhibitors for G2032R-ROS1, a rearrangement of the ROS1 protein which drives a subset of NSCLC for which is there is no treatment, using both IBDD and structure-based methods (SBDD). We evaluated our leads using molecular docking and discovered that the top lead generated through IBDD had a significantly higher binding affinity (-9.4 kcal/mol) than the top lead generated through SBDD (-8.2 kcal/mol), validating our approach. Leads were also evaluated using ADMET modeling, which showed that both molecules were nontoxic. We named our IBDD-generated lead Latrizanib and will continue with in vitro testing. IBDD has potential to disrupt the current pharmaceutical industry, speed up the drug discovery process, save billions of dollars spent every year on failed drugs, and increase financial accessibility to life-saving medications.
LATRIZANIB

Thinking Outside the Black Box: Competitive Inhibitor Targeting Crizotinib-Resistant G2032R-ROS1 Rearrangement using Novel Approach to *De Novo* Drug Discovery

Nora Sun[1], Ritvik Viniak[1], Cindy Ho[1]

[1]Walter Payton College Preparatory High School, Chicago, IL

INTRODUCTION

Deep Learning Applications in Drug Design

Traditional drug discovery processes were large-scale, costly, and time-consuming operations primarily driven by serendipity. The advent of modern drug discovery methods and resources such as virtual high-throughput screening (HTS), fragment-based screening (FBS), small molecule databases, and molecular docking drastically transformed the pharmaceutical industry. [1] Significant labor- and cost-intensive components of the discovery process can now be performed using software modules designed to evaluate a specific aspect of drug design. A sample in silico drug discovery pathway may start with the creation of the protein structure using homology modeling, followed by large-scale, CPU-intensive screening of biologically active molecules. Candidates are narrowed through pharmacophore or docking models and modeled for ADMET (absorption, distribution, metabolism, excretion, and toxicology). [2] Finally, a few dozen remaining candidates are tested using assays in vitro. These tools have been frequently implemented in the process of discovering drug leads in the past decade. [3] [4] The success of many of these programs is owed to a machine learning process called deep learning.

Recently, deep learning has been introduced to the world of drug design. Deep learning (DL), a more complex form of artificial neural networks (ANNs), mimics the human brain and has numerous applications ranging from image recognition to playing chess. DL is especially powerful at recognizing hidden patterns and delineating complex relationships. [5] Early ANNs consisted of only 3 layers (one input, one hidden, and one output) [6] and struggled with overfitting and large computational tasks. GPU acceleration and modified networks which optimized the weight-updating process for layers, transfer functions, and other components eventually allowed DL to become extremely powerful. [7]

DL has been applied to numerous areas of drug development including drug-target interaction prediction (DTI), molecule generation, and ADMET prediction. Notable work related to DL's application in drug discovery include Wang and Zeng's DTI-model using first-generation convolutional networks (CNNs) and Google's work on DL-based quantitative structure-activity relationship (QSAR) modeling. [8, 9] Since then, DL has rapidly replaced traditional machine learning (ML) in popularity on the drug development frontier owing to its ability to construct complex, nuanced approximation functions that generalize successfully. [10] Another key advantage of DL is its ability to create an accurate function without large amounts of data.

Despite the integration of DL in drug development, the productivity of research and development (R&D) continues to decline rapidly. [11] Accounting for the costs of failed trials, the capital investment to bring a single drug to market is approximately 4 billion dollars, with more expensive drugs such as nervous system agents costing upwards of 10 billion dollars. Additionally, despite improvements in silico technology, drug development costs have increased over the past decade, and the number of Phase I drugs that get approved is still only 10%. [12, 13] The rising cost of drug R&D can be attributed to numerous sources including the need for more complex drugs and the growing cost of human capital, but it makes it clear that, while current DL techniques may have improved the efficiency and cost of drug discovery, a more sophisticated approach which can increase lead success rates and decrease development cost is needed. [14]

De Novo Drug Design

The primary disadvantage of current in silico drug discovery is that strategies for finding leads typically rely on screening of virtual libraries with active reagents. While these libraries may contain millions of molecules, they are still finite and filtering often comes at a high computational cost. This strategy will always be limited to only discovered active reagents, and, even if it were not, it is presently infeasible to search all of predefined chemical space. [15] However, virtual screening (VS) remains the most popular strategy as it is the most well studied. Numerous machine learning methods have been applied to VS, ranging from sophisticated ANNs to simple k-nearest neighbors algorithms. [16]

De novo drug design is an underexplored strategy that employs DL or other computational methods to generate novel molecules. [17] One de novo drug design approach uses evolutionary or reinforcement learning algorithms to [18] and can generate molecules in text-based formats, such as SMILES, where molecules are represented by sequences of letters, and letters are generated based on the previous letters. However, this strategy is computationally intensive and time-consuming; additionally, multiple canonical molecules can be represented by a single SMILES. [19] More sophisticated multi-dimensional representations of molecules have been developed such as the work of Kang and Cho, who created graphical representation of molecules to replace SMILES. [20, 21] General adversarial networks (GANs), models which alter attributes of existing molecules, actor-critic models, and numerous other approaches have also been applied to de novo design. [22, 23]

However, many of these approaches face a similar limitation to VS. Most algorithms generate molecules based largely on structure and presence of functional groups. This is especially true with SMILES and models which alter attributes of existing molecules. While this is a reasonable approach because ligand structure plays a crucial role in ligand properties, this approach fails to optimize the quantity and quality of potent molecules discovered because it drives the algorithm towards a local minimum in the gradient descent that is in no way the guaranteed absolute minima. In other words, the SMILES, graphs, or other representations of molecules may converge at an optimal point, but this may not be the global optimal, and there could be a more optimal point that is completely overlooked because the algorithm favored a different type of structure given the training set. This problem stems from the "black box" of deep learning, a term used to describe the minimally understood and sometimes illogical processes which occur within the deep learning algorithms.

To avoid this pitfall, we employ an alternative, two-step approach that uses deep learning that 1) optimizes the binding energy of protein-ligand bonding interactions between the ligand and the protein (instead of the structure of the ligand), and 2) searches for the functional groups which can produce these desired interactions while performing well on ADMET (absorption, distribution, metabolism, excretion, toxicity). This process is termed as "Interaction-Based Drug Design" (IBDD) in contrast to conventional Structure-Based Drug Design (SBDD). To show that our approach designs more accurate molecules than traditionally screening, docking, or pharmacophore methods, we design a novel inhibitor for the crizotinib-resistant G2032R-ROS1 rearrangement which drives oncogenesis in NSCLC, for which no FDA approved drugs exist. We compare this inhibitor to one designed through conventional computational drug design methods.

Targeting Crizotinib-Resistant G2032R-ROS1 Rearrangement

In this study, we investigate a novel method using deep learning for lead optimization in potency, [illegible] to target the correct protein) and ADMET. We test this method on a Ros proto-oncogene1 rearrangement (ROS1). ROS1 is an orphan receptor tyrosine kinase (RTK) that activates downstream signaling pathways related to cell differentiation, proliferation, growth and survival such as the PI3 kinase-mTOR signaling pathway, which is activated by ROS1 phosphorylation of PTPN11. [24] However, ROS1 mutations lead to activation of the ROS1 tyrosine kinase, driving uncontrolled cell proliferation and oncogenesis in 1-2% of non-small cell lung cancer (NSCLC) cases. Crizotinib, an ATP-competitive tyrosine inhibitor, is the current first-line treatment for ROS1+ NSCLC. [25] Despite crizotinib's potency, various ROS1 resistance mechanisms have invariably developed. [26] ROS1 has acquired rearrangements which result in amino acid substitutions that induce tyrosine kinase inhibitor (TKI)-resistance. The most common ROS1 mutations are G2032R (41%), D2033N (6%), and S1986F (6%). [27] No treatment exists for NSCLC subset of TKI-resistant ROS1-G2032R, which accounts for roughly 2,000 cases of lung cancer annually. [28]

The binding site of G2032R is located on amino acid 1980, while the active site is located on amino acid 2079. During fusion, the kinase site is preserved, making it an optimal target for drugs such as crizotinib. However, the G2032R mutation mutates a small glycine in a beta turn to a large and bulky arginine. Though the mutation occurs in a disordered region, meaning it is unlikely to have a large impact on

the structure of the protein, the side chain of the arginine obstructs the active site where crizotinib previously inhibited the protein. [29]

Figure 1. G2032R mutation prevents crizotinib from inhibiting ROS1. Visualizations are generated using UCSF Chimera and PyMol. 1a) Wild type ROS1 in complex with crizotinib. 1b) Crizotinib cannot bond effectively with rearranged ROS1 due to the bulky arginine. 1c) Closer display of arginine's large side chain obstructing crizotinib's bonding.

Figure 2. Pharmacophore modeling using LigandScout and PharmaGist. Identified pharmacophores were used to screen [illegible]

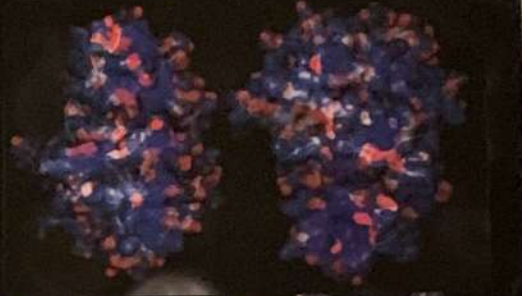


Figure 3. Right: The MaSIF technique was used to estimate the potential for various interactions on the surface of the protein. The active site is circled on the left protein.

Figure 4. Below: 4a) Depictions of the latrizanib molecule (blue) in complex with mutated ROS1 (grey spheres) in the binding pocket. 4b) Magnified depiction of latrizanib molecule in binding pocket. The mutated side chain arginine 2032 is shown in spherical form; it no longer obstructs the binding of latrizanib as it did crizotinib. 4c) The latrizanib molecule before binding.

METHODS

Pharmacophores and molecular docking, standard in silico procedures for lead identification, were used to identify a competitive inhibitor for crizotinib-resistant G2032R-ROS1 rearrangement which drives oncogenesis in NSCLC. Then, a second lead molecule is identified using novel IBDD methods proposed in this study. The potency and ADMET properties of each lead molecules will be measured through ADMET and molecular docking to test whether the IBDD lead performs better than current methods.

Structure-Based In Silico Development Methods

Binding Site Visualization. PyMol 2.4.0 and UCSF Chimera 1.14 desktop applications were used for visualization and analysis of the binding site. [30, 31]

Homology-based Modeling of G2032R Mutation. No crystal structure of ROS1 with G2032R mutation existed in the Protein Database (PDB) as of October 6, 2020. Biskit (URL: http://biskit.pasteur.fr/) Version 2.4 was used to determine the structure of ROS1 G2032R with automated homology modeling. The sequence of unmutated ROS1 was used as the homologous sequence.

Pharmacophore-based Search. The open-source interfaces PharmaGist (URL: https://bioinfo3d.cs.tau.ac.il/PharmaGist/) and ZincPharmer (URL: http://zincpharmer.csb.pitt.edu/) were used to identify pharmacophores and search the ZINC database, which contains 176 million conformers of 18.3 million active compounds. [32]

Molecular Docking. Autodock Vina 4 was used to perform molecular docking on the highest scoring molecules from the pharmacophore based search. [33]

ADMET Modeling. SwissADME (URL: www.swissadme.ch) is used to model ADMET properties of top-performing molecules from docking. [34]

Novel Interaction-Based Drug Design

Interaction Fingerprints: Protein Surface Feature Identification. Predicting potential interactions between a protein and ligand purely from the protein structure remains a large challenge in drug development. While some models, such as pharmacophores, have attempted to describe molecular features which are important to protein-ligand binding, very few pre-existing methods capable of predicting the types of interactions various sites on proteins are capable of. For the purposes of this paper, we have adopted Molecular Surface Interaction Fingerprints (MaSIF), pioneered by Gainza et. al, as the closest existing algorithm which completes our goal. MaSIF stands for molecular surface interaction fingerprinting and utilizes geometric deep learning, a method which consistently outperformed handcrafted feature extraction, to describe protein surface interaction fingerprints.

MaSIF carves out overlapping radial patches on the protein surface. Each path has a fixed geodesic (a Riemannian manifold describing the shortest path between two points) radius. Each patch has pre-computed geometric and chemical properties mapped on the protein surface; these properties are translated into numerical vector descriptors which are used to train end-to-end networks. [34] Once patches are selected, two geometric features (shape index and distance-dependant curvature) and three chemical features (hydropathy index, continuum electrostatics and the location of free electrons and proton donors) are identified for each patch. Within these patches, vertices are assigned geodesic polar coordinates. With this framework to describe surface patches, the information is further processed by various neural networks. [35]

The Github open-source code project by "MaSIF- Molecular Surface Interaction Fingerprints: Geometric deep learning to decipher patterns in protein molecular surfaces" by Gainza et. al was employed for this step.

Following identification of features using MaSIF, a corresponding ligand needs to be generated that maximizes the binding strength between the ligand and the protein based on the features identified.

De Novo Drug Design Using Fragment-growing Method. Using protein surface features and potential interactions extracted from MaSIF, a novel inhibitor can be generated the optimizes protein-ligand bonding strength. In this step, MaSIF properties are trained by the model to be incorporated into bonding forces such as hydrophobic interactions, like van der Waals forces, or polar interactions. The de novo drug design method, fragment-growing, is employed for this step.

A novel ligand is designed by matching fragments with certain properties at coordinates identified though MaSIF. Initial fragments are selected from sets of fragments matching the ideal properties (e.g. one that would allow for strong polar interactions) and are subsequently bonded to the next fragment. The model uses layers of networks to determine the molecule with the strongest interactions, i.e. the highest binding energy. [36, 37] Utilizing pre-existing fragments leads to high diversity among generated molecules and minimizes the likelihood of chemically infeasible or highly toxic molecules. Additionally, as chemical space is explored through different sets of fragments, parameters are experimentally determined to arrive at the molecule which provides the strongest interactions. Because this model is based off the strength of interactions instead of the physical structure of the molecule, aforementioned gradient descent bias is thwarted. [38]

Code from the Github open-source projects "FragPELE" by Lopez et. al for fragment-growing drug discovery and "ProtoCaller" for protein-ligand binding free energy calculations were merged with the features data collected from the MaSIF process to grow the molecule with the highest binding energy. [39]

RESULTS

Structure-Based *In Silico* Development Methods

During pharmacophore-based development, the ZINC database was narrowed to 5 molecules (ZINC13728774, ZINC04659596, ZINC02985419, ZINC08984285, ZINC15800138). All 5 molecules passed ADMET testing, signifying that they had acceptable ADMET qualities and have potential to be made into a drug. (Unfortunately, testing was limited primarily to toxicity properties due to lack of access.) In molecular docking, the molecules scored binding affinities of -7.2, -7.6, -8.9, -7.8, and -8.2 kcal/mol.

Structure-Based *In Silico* Development Methods

During MaSIF, the primary bond potentials identified were hydrogen bonding and hydrophobic interactions. A molecule was generated which optimized interactions between the protein and the ligand. The molecule passed ADMET testing, signifying that they had acceptable ADMET qualities and have potential to be made into a drug. (Unfortunately, testing was primarily limited to toxicity properties due to lack of access.) In molecular docking, the molecule scored a binding affinity of -9.4 kcal/mol, which is not only significantly higher than the next highest trial determined using SBDD (-8.2 kcal/mol), but also higher than the binding affinity of crizotinib in unmutated ROS1 (-8.6 kcal/mol). The binding was verified in PyMOL, which showed that the molecule ($C_{21}N_5H_{26}OF$) 1O1 binds to the correct active site (though at a slightly different angle) and is not obstructed by the mutated arginine residue at 2032.

The Latrizanib Treatment

In this study, we proved that IBDD design outperformed SBDD. We name the de novo $C_{21}N_5H_{26}F$ 1O1 molecule latrizanib and hope to conduct in vitro studies to verify our results.

DISCUSSION

Analysis

In this study, we discussed a novel method of in silico drug development. While previous computational drug development methods have relied largely on the structure of the protein or ligand, our method focuses on the potential interactions this protein may have with the ligand. MaSIF, a geometric deep learning method that describes protein surface interaction fingerprints, determine the interactions the protein is looking for at a certain site. Then, fragment-growing is used to create a minimally toxic ligand that maximizes protein-ligand binding affinity. In our study, IBDD methods outperformed SBDD methods. IBDD methods were able to identify a new approach to binding to the active site which SBDD methods could not identify, and the de novo molecule generated by IBDD performed significantly better in molecular docking than all SBDD generated molecules.

Impact

Not only do we discuss a novel, high-potency inhibitor for ROS1-G2032R, a subset of NSCLC that currently has no treatment, we also pioneered a novel approach for de novo drug synthesis. [28] We hope our work will encourage drug development to shift away from focusing on structure and more directly towards the interactions these structures enable. As drugs grow more and more expensive and unattainable for many patients because multi-million-dollar experiments fail to produce a drug on the market, it is imperative to turn to in silico development. With refinement and application of our methodology, IBDD could save billions of dollars in failed trials and speed up the drug development process by more than 10x. [12, 13]

Future of Latrizanib

Next, the latrizanib molecule will undergo preclinical in vitro and in vivo testing and clinical trials. If successful, the drug will be administered orally as a pill. Specific dosage will be determined through clinical trials and testing. It will be incorporated into a typical NSCLC chemotherapy regimen such as with vinorelbine.

The drug will be metabolized by CYP3A4 or CYP3A5 from the liver in a pathway similar to Lapatinib. Upon diffusing into the bloodstream, it will reach the ROS1 active site in the tumor region and inhibit ROS1, effectively ceasing tumor cell production and ultimately causing tumor cell death.

Future of Latrizanib

We envision that IBDD drug design can be implemented into the drug design process in the form of a software package that can be trained and improved over time. IBDD will fit into the beginning of the R&D process in drug development, allowing drug developers to find more potent and less toxic leads significantly faster and at much lower costs.

IBDD is capable of disrupting the drug development and pharmaceutical industry in numerous ways. While many companies already incorporate some computation into their processes, IBDD can shift lead discovery to a fully in silico process. There will be a decrease in wet lab or non-computational positions such as experimental medicinal chemistry. However, there will be an increase in demand for workers with strong computer science and software backgrounds such as computational chemists and biochemists in the pharmaceutical industry.

The average cost of a single dose of a cancer drug is more than $100,000 per person per year. This can partially be attributed to the high cost of high throughput drug screening and other in vivo testing activities, which are hundreds of thousands of dollars. More importantly, however, drugs are priced high to account for the cost of failed trials and testing for medications which are never able to make it to the market. By using our IBDD approach, the amount of funding spent on failed drugs will be minimalized so that drug costs can be significantly lowered without curbing innovation. We estimate that IBDD-developed drugs such as Latrizanib can eventually lower drug costs to $10,000 and increase drug availability.

ABSTRACT

Traditional lead discovery is very expensive and has a high rate of failure, driving up drug prices. However, the shift towards in silico lead discovery has been hindered by in silico methods' high error rate. We explored a novel approach to de novo lead discovery that relies primarily on potential interactions between protein and ligand, termed interaction based drug discovery (IBDD), instead of protein and ligand chemical structure. IBDD is achieved by using molecular surface interaction fingerprinting (MaSIF) to identify potential interactions on the protein's surface and fragment growing to build leads from these fingerprints. In this study, we identified inhibitors for G2032R-ROS1, a rearrangement of the ROS1 protein which drives a subset of NSCLC for which is there is no treatment, using both IBDD and structure-based methods (SBDD). We evaluated our leads using molecular docking and discovered that the top lead generated through IBDD had a significantly higher binding affinity (-9.4 kcal/mol) than the top lead generated through SBDD (-8.2 kcal/mol), validating our approach. Leads were also evaluated using ADMET modeling, which showed that both molecules were nontoxic. We named our IBDD-generated lead Latrizanib and will continue with in vitro testing. IBDD has potential to disrupt the current pharmaceutical industry, speed up the drug discovery process, save billions of dollars spent every year on failed drugs, and increase financial accessibility to life-saving medications.