FABAD J. Pharm. Sci., 49, 3, 583-602, 2024
Doi: 10.55262/fabadeczacilik.1411661
REVIEW ARTICLE

The Rise of Artificial Intelligence in Pharma: Shaping the Future of Drug Discovery

Ayca DEDEOGLU-ERDOGAN[*°], Armanc MAT[**], Enise Ece GURDAL[***], Meric KOKSAL[****]

The Rise of Artificial Intelligence in Pharma: Shaping the Future of Drug Discovery

SUMMARY

Drug discovery as an important scientific area that serves human health, requires continuous advancement for improved quality of life and survival rates. However, drug discovery is a long and expensive process. The studies aimed at dealing with these problems have enabled to combination of artificial intelligence (AI) with drug development stages. For every step of the R&D process, AI plays a vital role in facilitating and accelerating the work. Firstly, AI methods (deep learning and convolutional neural networks) help predict the 3D structure of protein making it easier for the rational design of compounds to target a specific protein among other potential outcomes. After estimation of the protein structure of interest, it is also possible to determine the protein-ligand interactions by utilizing AI technologies like random forest. The other stage, namely finding the hit compounds is also possible through AI-assisted QSAR models such as deep neural networks. Besides, there are many AI methods (k-nearest neighbor and support vector machines) for ADMET prediction to optimize lead compounds. Finally, AI techniques also aid in choosing the most suitable synthesis plan. In the light of the latest advances, AI has become the focus of the pharmaceutical industry. However, despite the potential benefits of AI in drug discovery, several challenges must be considered including the availability of suitable data and bioethical issues. This article provides a comprehensive review of the benefits and applications of AI in various stages of drug discovery.

Key Words: *Artificial intelligence, drug discovery, machine learning, deep learning*

Yapay Zekanın Eczacılıkta Yükselişi: İlaç Keşfinin Geleceğini Şekillendirmek

ÖZ

İnsan sağlığına ve refahına hizmet eden önemli bir bilimsel alan olan ilaç keşfi, yaşam kalitesinin ve hayatta kalma oranlarının iyileştirilmesi için sürekli ilerlemeyi gerektirmektedir. Ancak ilaç keşfi uzun ve pahalı bir süreçtir. Bu sorunların üstesinden gelmeye yönelik çalışmalar, yapay zekanın ilaç geliştirme süreciyle birleştirilmesini sağlamıştır. Ar-Ge sürecinin her adımında yapay zeka, işi kolaylaştırma ve hızlandırma konusunda hayati bir rol oynar. Öncelikle, bazı yapay zeka yöntemleri (derin öğrenme ve evrişimsel sinir ağları), ilaç molekülünün belirli bir proteini hedeflemesini kolaylaştırmak için proteinin 3 boyutlu yapısını tahmin etmeye yardımcı olmaktadır. Protein yapısının modellenmesiyle beraber, rastgele orman gibi yapay zeka algoritmalarından yararlanılarak protein-ligand etkileşimlerinin belirlenmesi de mümkün olmaktadır. Takip eden aşamada, öncü bileşiklerin keşfi de derin sinir ağları gibi yapay zeka destekli QSAR modellerinin kullanılmasıyla sağlanmaktadır. Ayrıca, öncü bileşiği optimize etmek amacıyla ADMET tahminine yardımcı birçok yapay zeka yöntemi de (k-en yakın komşuluk, and destek vektör makineleri) bulunmaktadır. Bunun yanı sıra yapay zeka teknikleri en uygun sentez planının seçilmesinde de yol göstericidir. Güncel gelişmelerin ışığında yapay zeka, ilaç sektörünün odak noktası haline gelmiştir. Ancak, ilaç keşfinde yapay zekanın potansiyel faydalarına rağmen, uygun verilerin mevcudiyeti ve biyoetik konular da dahil olmak üzere dikkate alınması gereken çeşitli zorluklar da vardır. Bu makale, yapay zekanın ilaç keşfinin çeşitli aşamalarındaki yararları ve uygulamalarını kapsamlı bir şekilde incelemektedir.

Anahtar Kelimeler: *Yapay zeka, ilaç keşfi, makine öğrenmesi, derin öğrenme*

Received: 29.12.2023
Revised: 22.04.2024
Accepted: 24.09.2024

[*] ORCID: 0000-0001-5936-5328, Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Yeditepe University, Atasehir, 34755 Istanbul, Turkey
[**] ORCID: 0009-0001-8379-3987, Faculty of Pharmacy, Yeditepe University, Atasehir, 34755 Istanbul, Turkey
[***] ORCID: 0000-0003-1064-8639, Department of Organic Chemistry, Faculty of Chemistry, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany
Department of Nano-optics, Max Planck Institute for the Science of Light, 91058, Erlangen, Germany
[****] ORCID: 0000-0001-7662-9364, Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Yeditepe University, Atasehir, 34755 Istanbul, Turkey

[°] Corresponding Author; Ayca DEDEOGLU-ERDOGAN,
E-mail: elifaycadedeoglu@gmail.com

INTRODUCTION

The Pioneering Steps of Artificial Intelligence in Healthcare

The notion of incorporating machine-based intelligence into human daily life has been a longstanding concept, with roots in ancient myths and the creation of automatons in Chinese and Egyptian cultures (Lewis, 2014). Intelligent behavior, in this context, is generally defined as making decisions according to the information collected by the system. This concept eventually paved the way for the development of Artificial Intelligence (AI) in 1955, marking a significant milestone in the intersection of machines and intelligent decision-making. Subsequently, the integration of AI into various parts of the developing world gained momentum and its entrance into the pharmaceutical field occurred in the 70s with the introduction of Dendral at Stanford University (Lindsay *et al*. 1993). Regarded as one of the earliest applications of AI in the healthcare system, Dendral played a crucial role in identifying unknown organic molecules. It achieved this by analyzing mass spectra and using the information from chemical databases available at the time. Since then, AI has been applied in numerous areas of healthcare systems including personalized medicine, assisted diagnostics, health economics, drug discovery, and development. It is important to note that AI encompasses a spectrum of definitions and approaches, leading to a diverse range of solutions addressing industry challenges today.

Machine Learning And Deep Learning Models

In an AI concept, it is essential to cover key subfields such as solution searching, reasoning, and knowledge representation, with particular emphasis on the primary paradigm of machine learning (ML). ML is an evolving field of computational algorithms which are designed to imitate human intelligence by learning through the surrounding environment and experiences (Naqa and Murphy, 2015). k-nearest neighborneighbor (k-NN), support vector machines (SVM) and random forest (RF) are the main types of ML (Cortes and Vapnik, 1995, Lavecchia *et al*., 2013, Melville *et al.,* 2009). The k-NN algorithm, a form of instance-based learning or lazy learning, is a simple and intuitive method to predict the class, property, or rank of a molecule based on the nearest training examples in the feature space. It has been used for envisaging activities of various compounds such as anti-convulsants, dopamine D1 antagonists, protein kinase inhibitors, psychoactive cannabinoids, steroids, anti-inflammatory and anticancer drugs, and estrogen receptor agonists (Lavecchia, 2015). Support vector machines (SVMs), as supervised machine learning algorithms, have gained popularity in drug discovery applications for tasks like compound classification and property predictions for novel active compounds. In a study applying SVM to pharmacokinetic (PK) modeling, Doniger *et al.* worked on 179 central nervous system (CNS) active compounds and 145 inactive molecules for predicting blood-brain barrier (BBB)penetration. Using 30 tests, the AI model was performed with 81.5% success (Heikamp and Bajorath, 2014, Doniger *et al.*, 2002). Random Forest (RF) is a supervised learning method to be applied to classification and some problems. The method involves tree predictors that each tree depends on the values of a random vector independently and with the same layout for each of the generated vectors (Breiman, 2001). The optimal selection of chemical features (molecular descriptors) is an important act to take before the research for the efficient application of AI techniques in virtual screening for identification of bioactive molecules in drug discovery. This selection plays a role in the accuracy of affinity prediction. RF-based approaches automatically select molecular descriptors of training data for ligands of kinases, nuclear hormone receptors, and other enzymes (Cano *et al*., 2017).

Deep learning (DL), a subfield of ML, involves artificial neural networks (ANNs), and provides advantages compared to statistical modeling. One

notable advantage of DL is that it does not require rigidly structured experimental designs, instead, it can map functions using historical or incomplete data. ANNs are good recognizers of patterns and robust classifiers, with the ability to generate when making decisions based on imprecise input data (Cheng and Sutariya, 2012). ANNs have many application areas in drug design such as pharmacokinetics, neurodegenerative diseases, cardiovascular diseases, infectious and microbial diseases, immunology and virology, medical diagnosis, cosmetics and dermatology, proteomics-genomics etc. (Dobchev and Karelson, 2016). The computational model inspired in the natural neurons, ANN, covers the interconnected and sophisticated computing factors to process information and solve problems. ANNs comprise different types, such as multilayer perceptron (MLP) networks, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). The MLP network is used for pattern recognition, optimization aids, process identification, and controls. RNNs can memorize and store information. CNNs have used in biological system modeling, processing complex brain functions, pattern recognition, and sophisticated signal processing (Paul *et al.* 2021). Examples of method domains of AI are summarized in Figure 1.

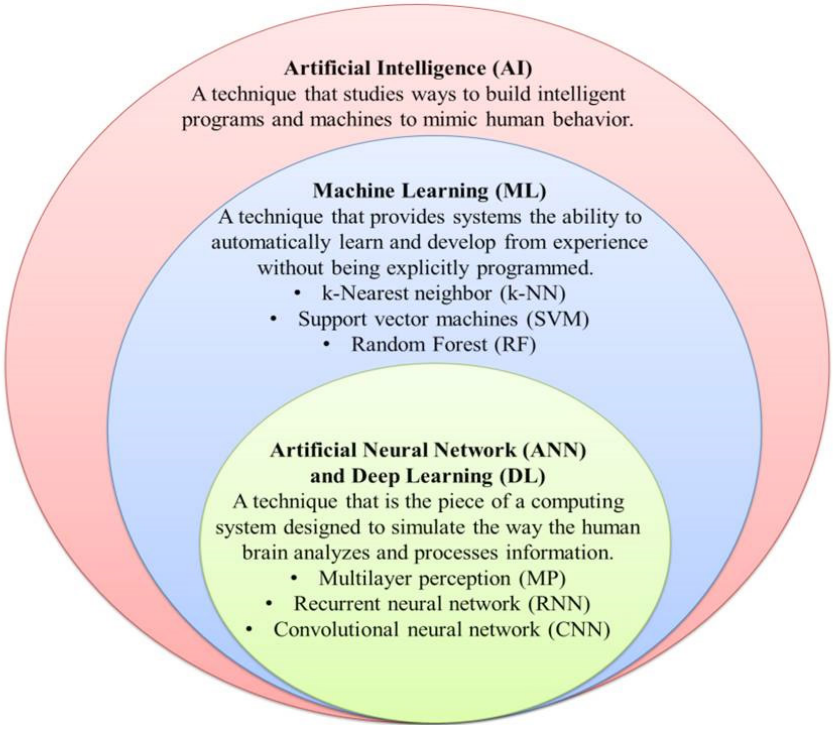

Figure 1. Method domains of AI

Drug Discovery Process

The drug discovery process is a complex and long pathway that consists of considerable steps. The initial phase involves evaluating of three-dimensional (3D) structure of the target protein and examining protein-protein/drug interactions. Following target identification, potential hit compounds targeted on selected protein are chosen by using computer-aided drug design systems such as molecular docking, virtual screening, and drug repurposing (Rao and Srinivas, 2011). Once identified, the selected compounds are synthesized and then the compounds undergo *in vitro* and *in vivo* assays to determine their activities, toxicities, and pharmacokinetic properties. Moreover, quantitative structure–activity relationship (QSAR) and *de novo* studies are utilized in combination with the

activity data for lead identification and optimization steps. After these steps, the drug candidates are selected for preclinical and clinical studies for more in-depth analysis (Schneider and Fechner, 2005). The drugs that successfully pass all the clinical phases are reviewed by FDA or any other autority to be ready for the market (Chan *et al*., 2019). The drug development process is visually summarized in Figure 2.

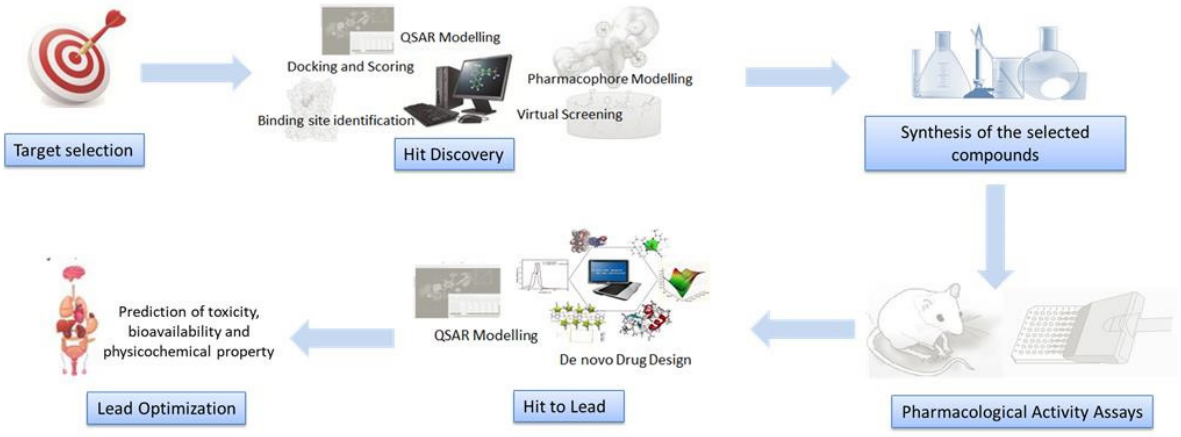

Figure 2. AI-aided drug development process

Drug development is an extended, complex, and costly process, marked by a considerable degree of uncertainty regarding the success of a drug (Fuloria *et al*., 2013). The research and development (R&D) process requires analyzing a large amount of data for hypothesis and compound identification. It necessitates the integration of datasets with the targeted pathophysiology, extraction of pertinent data aligning with specific research goals, and establishing correlations between subjects for improvement. For every step of the R&D process, AI plays a pivotal role in streamlining and expediting the work.

To discover an effective agent, conducting error-free lab tests or clinical trials is imperative, as every error poses an obstacle and hinders the progress of the process. Therefore, developing an agent that can efficiently treat individuals with highly accurate results is paramount. The integration of advanced techniques, such as AI models, into the drug discovery process, holds the potential to mitigate human-based errors and save valuable time. Major biopharmaceutical companies are increasingly turning to artificial intelligence as a solution to the challenges in drug design. Pfizer, for instance, employs a machine learning system called IBM Watson to enhance the search for immuno-oncology drugs. Sanofi has agreed to utilize the artificial intelligence platform developed by the UK-based Exscientia to research the treatment of metabolic disorders. Numerous other biopharmaceutical companies have similar collaborations or internal AI programs, reflecting a growing consensus within the industry that AI can offer solutions to critical challenges in drug development (Fleming, 2018).

Target Identification

The overexpression of numerous proteins is the fundamental reason for the progress of many diseases. The successful therapeutic approach for these diseases is design and development of drug molecules that target the overexpressed proteins. Therefore, it is essential to predict the structure of the protein to design the drug molecule that selectively interacts with it (Madhukar *et al*., 2017). Structure-based drug design, which involves examining the three-dimensional (3D) structures of proteins, is a valuable strategy for identifying active small molecules targeting protein of interest. However, measuring the

3D structure of proteins is a time-consuming and expensive process, prompting the development of algorithms to predict protein structures to overcome these problems. Although the sequence knowledge of most proteins is currently available, achieving precise *de novo* prediction of their 3D structures remains a significant challenge. Recently, the powerful capability of feature extraction has enabled the utilization of deep learning technologies in predicting the secondary structure, backbone torsion angle, and residue contacts of proteins (Spencer *et al.*, 2014, Li *et al.*, 2017). The use of deep learning modules, particularly convolutional neural networks (CNN), is advantageous in addressing the overfitting problem in protein structure prediction (Wang and Zhang, 2017). Notably, the AlphaFold2 model has emerged as a highly accurate tool for predicting protein folding, as evidenced by the impressive results in the CASP14 assessment. This model is based on the neural network algorithm that incorporates multi-sequence alignments and pairwise features, refining predictions iteratively (Jumper *et al.*, 2021). The landscape of AI-based computational tools for target identification is further detailed in Table 1.

In research, Jumper and colleagues presented a computational method that could predict protein structures with near-experimental accuracy. Their created neural network AlphaFold was submitted to CASP14. AlphaFold structures outperformed rival approaches in CASP14 by a significant margin. The next best performing method had a median backbone accuracy of 2.8 Å r.m.s.d.95, while AlphaFold structures had a median backbone accuracy of 0.96 Å r.m.s.d.95 (Figure 3a). A carbon atom's width is around 1.4 Å, which can be used as a benchmark for precision. When the backbone is highly precise, in addition to extremely accurate domain structures (Figure 3b). AlphaFold may generate highly accurate side chains (Figure 3c) . This significantly outperforms template-based approaches, even in the presence of strong templates. AlphaFold's all-atom accuracy was 1.5 Å r.m.s.d.95, while the best alternative method's all-atom accuracy was 3.5 Å r.m.s.d.95. With precise domains and domain-packing, our approaches scale to exceedingly long proteins (Figure 3d). Lastly, the model's ability to produce accurate, per-residue dependability estimations should allow users to utilize these predictions with confidence (Jumper *et al.*, 2021).

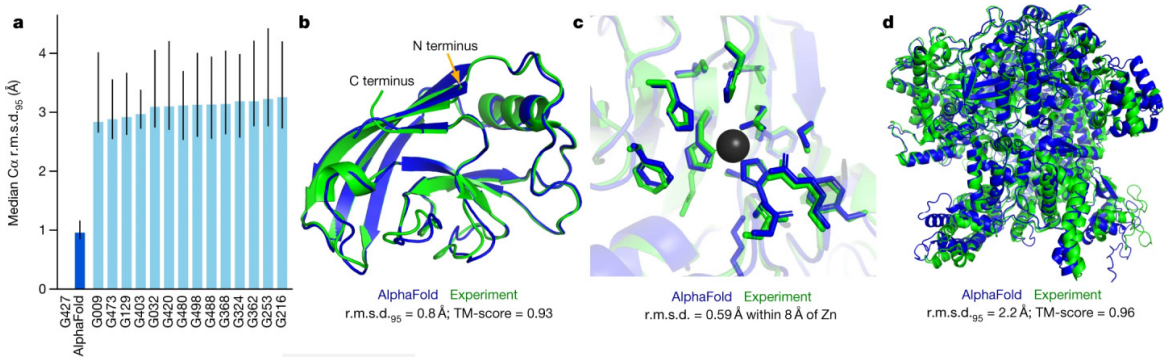

Figure 3. High prediction accuracy of AlphaFold structure. **a**, AlphaFold's performance on the CASP14 dataset (*n* = **87 protein domains)** in comparison to the top-15 entries (out of 146 entries), group numbers match the numbers given to participants by CASP. The information is provided by the median and its 95% confidence interval, which are calculated using 10,000 bootstrap samples. **b**, the real (experimental) structure (green) and the AlphaFold prediction of CASP14 target T1049 (PDB 6Y4F, blue) are comparisons. Four residues in the crystal structure's C terminus are not shown because they are B-factor outliers. **c**, CASP14 target T1056 (PDB 6YJ1). An illustration of a zinc-binding site that is accurately predicted (while AlphaFold does not specifically predict the zinc ion, it does have correct side chains). **d**, CASP target T1044, a 2,180-residue single (PDB 6VR4) chain was predicted with correct domain packing. The figure is reproduced from Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2. Copyright © 2021, The Author(s). The figure was cropped from the original one retaining the data provided by authors.

A study used AI-generative models to create a series of hinge cores based on the binding posture of a reported chemical (GLPG-3970, 3) with the AlphaFold protein structure (Chemistry42). A hit molecule targeting SIK2 was produced using a new scaffold following molecular docking, manufacturing, and biological assessment. Compound **8g** was found through additional SAR investigation to have better efficacy against SIK2 than the previously reported inhibitors (Figure 4). *In vitro* studies also confirmed the *in silico* studies, proving that this compound has high activity, good ADMET profiles, and great selectivity over other AMPK kinases. As a result, an artificial intelligence method for finding new and selective kinase inhibitors is offered by this work (*Zhu et al.*, 2023).

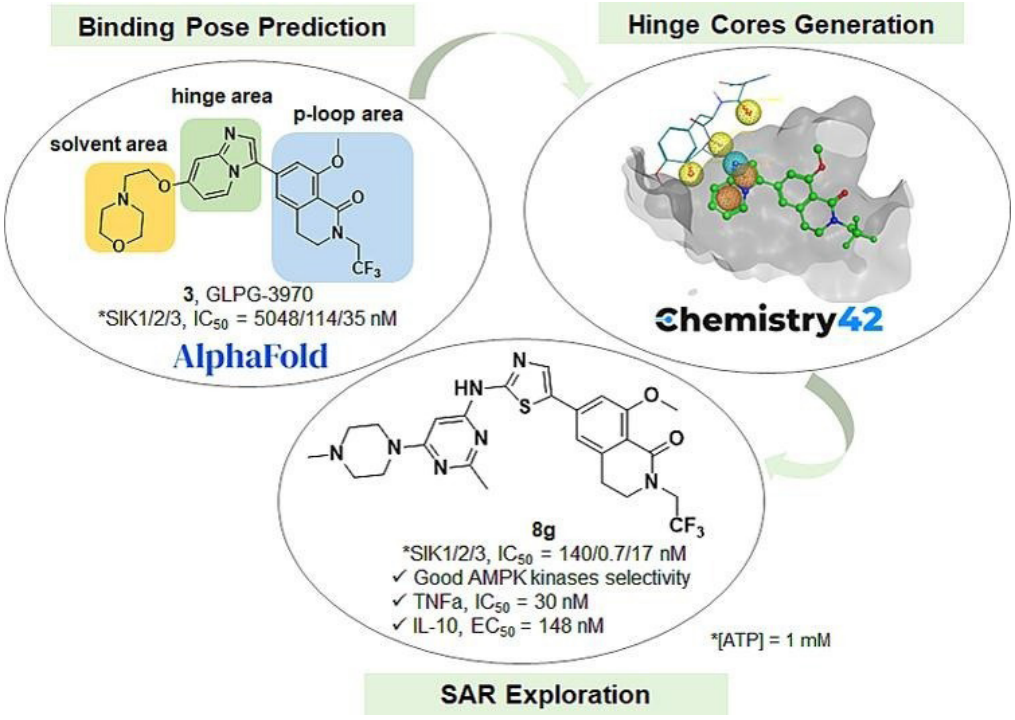

Figure 4. Design of compound 8g as AMPK kinase inhibitor (Zhu *et al.,* 2023)

Table 1. AI-Based computational tools for target identification

Tools	Description (available websites)	References
AlphaFold2	An AI system developed by DeepMind to predict a protein's 3D structure from its amino acid sequence. (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb)	Jumper *et al.*, 2021
RosettaFold	A three tract neural network, learning the patterns in sequences, and protein's amino acid interactions with one another, and predicts a protein's three-dimensional structure.	Baek *et al.*, 2021
DeepFragLib	Protein-specific fragment library built using deep neural networks.	Wang *et al.*, 2019
ProteinNet	A standardized data set for machine learning of protein structure to provide protein sequences, structures, MSAs, PSSMs, and standardized training/validation/test splits.	Cao *et al.*, 2012b

MSA: multiple sequence alignments, PSSM: position-specific scoring matrices

Hit Discovery

Virtual screening (VS) is one of the primary computational methods in drug discovery to identify bioactive molecules capable of binding target proteins. It is an efficient method that is used in early drug development to eliminate compounds that cannot target the protein of interest and to identify new hits (Lavecchia and Giovanni, 2013). As a result, virtual screening has become an indispensable tool in overcoming challenges associated with high costs and low success rates in drug discovery. Virtual screening methods are broadly categorized into two groups: structure-based drug design (SBDD) and ligand-based drug design (LBDD). SBDD relies on understanding the possible interactions between ligands and the structurally resolved target protein. On the other hand, LBDD focuses on assessing the similarity of a designed compound to known bioactive agents. SBDD needs 3D structural knowledge of the target protein (Dror *et al.*, 2004). The most used technique for SBDD, molecular docking, provides the prediction of the binding pose of the ligand in the target protein and determines its binding affinity (Blundell, 2019). In molecular docking techniques, many possible ligand poses have relied on the target protein, and the ligands are ordered by a scoring function (SF). Molecular docking programs implement a search algorithm in which the conformation of the ligand is determined recursively up to the convergence to the minimum energy is achieved. Finally, as the sum of the electrostatic and Van der Waals energies, an affinity scoring function, ΔG U total i kcal/mol, is used to rank the candidate postures. The forces that propel these particular interactions in biological systems are directed at complementarities between the ligand or substrate and the binding site surfaces in terms of electrostatics and shape (Pagadala *et al.*, 2017). Consequently, an increase in protein-ligand binding and structural data makes it possible to identify the protein-ligand interactions by using AI technology, which provides progress in SBVS (Table 2).

Recently, many researchers have utilized molecular docking techniques to identify the possible interactions between compounds and target proteins. For example, in a study, to combat viruses a new class of compounds was introduced and to assess their antiviral efficacy against the major protease M pro of SARS-CoV-2 (2019-nCoV) the Auto Dock Vina program was used. As a result, when compared to lopinavir as a reference drug, three compounds had the most promising antiviral efficacy against SARS-CoV-2. The findings highlight the consistency of the in vitro and in silico studies (Alamshany *et al.*, 2023).

Table 2. AI-Based computational tools for hit discovery

Tools	Description (available websites)	Reference
RepCOOL	A novel network-based method for drug repositioning.	Fahimian *et al.*,2020
DeepConv-DTI	A deep learning method to predict drug-target interaction. (https://github.com/GIST-CSBL/DeepConv-DTI)	Lee *et al.*, 2019
DeepH-DTA	A deep learning method to predict drug-target interaction. (https://github.com/Hawash-AI/deepH-DT)	Abdel-Basset *et al.*, 2020
DeepPurpose	Provides the library for drug-target prediction based on deep learning.	Huang *et al.*, 2020
AutoDock	Provides the prediction of how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. (https://autodock.scripps.edu/)	Österberg *et al.*, 2002
MOE	A drug discovery software platform that integrates visualization, modeling, and simulations, as well as methodology development, in one package. (https://www.chemcomp.com)	Reynolds et al., 2010
GLIDE	A molecular modeling software developed by Schrödinger, for docking of small molecules into proteins and other biopolymers.	Pagadala *et al.*, 2017

Recently, new categories of machine and deep learning-based scoring functions (SFs) have been introduced to discern the relationship between interaction terms for predicting binding affinity. ML methods such as RF and SVM are leveraged to enhance the efficacy of SFs (Ballester, 2019; Ain *et al.,* 2015; Shen *et al*., 2019; Coley *et al*., 2020; Li *et al.,* 2020). These approaches introduce non-linear relationships between individual energy terms and binding affinity, notably improving screening and scoring capabilities. For instance, Wang and Zhang incorporated a ΔvinaRF parameterization correction technique, integrating RF with AutoDock scoring function, demonstrating excellent performance compared to GlideScore XP (Repasky *et al.*, 2012).

Traditional ML methods face limitations in manual recognition and feature extraction, hindering large-scale applications. The emergence of deep learning (DL) methods addresses this challenge. Capitalizing on the success of CNN in image processing, this technique is employed to extract features from protein-ligand interaction maps for predicting protein-ligand affinity. Jimenez *et al* utilized a 3D graph CNN model that provides a good relationship with experimental data in the datasets to show predictive binding affinities (Jimenez *et al*., 2018). Pereira *et al*. established a deep convolutional neural networks method called DeepVS that gets the outcomes of MD as the input of DCNN, and can automatically learn and extract relevant features from the basic data (Pereira *et al*., 2016, Jiang *et al.*, 2018).

Hit To Lead Optimization

In the process of lead optimization, potent lead compounds can be found by analyzing and predicting the activity of a series of drug analogs. The quantitative structure-activity relationship (QSAR) models for virtual screening are derived by the standard ligand-based drug design to find the potent candidates from a series of hits compounds by prediction of pharmacological activity. QSAR mainly refers to the use of mathematical methods for examining the quantitative mapping relationship of the structural or physicochemical properties of compounds with their pharmacological activities. By screening the molecular database, QSAR method automatically chooses the most promising compounds for synthesis and analysis. It saves time and money by reducing the blindness of the experiment and accelerating the drug development process with the desired pharmacological activity. The process of QSAR method consists of data collection, data selection, generating molecular descriptors, the establishment of a mathematical model, interpretation, and application of models. With the development of ML techniques, AI models are used in QSAR research to construct mathematical models of the relationship between chemical structure and pharmacological activity (Zhong *et al*., 2018, Dobchev *et al*., 2014, Dudek *et al*., 2006, Ning and Karypis, 2011). Neural networks (NNs) method was introduced to QSAR analysis by Aoyama *et al*. in 1990. Various traditional ML methods, such as RF and SVM, have also been widely utilized to construct QSAR models (Aoyama *et al*., 1990). In recent years, DL methods have been implemented to QSAR modeling due to the ability of dealing with various chemical characters and the merit of extracting features automatically (Ghasemi *et al*., 2018) (Table 3).

Numerous research on the use of AI-assisted QSAR models, such as RF and DNN, to identify the hit compounds have been published in recent years. To show that RF and DNN were better in hit prediction efficiency, Tsou LK and colleagues conducted comparison tests between DNN and other ligand-based virtual screening (LBVS) techniques. Several triple-negative breast cancer (TNBC) inhibitors were identified as strong hits by DNN screening of the 165,000-compound database. Their findings demonstrate the potential of DNN as an effective hit prediction module and offer experimental proof that machine learning is capable of identifying strong hits *in silico* from a small training set (Figure 5) (Tsou *et al*., 2020).

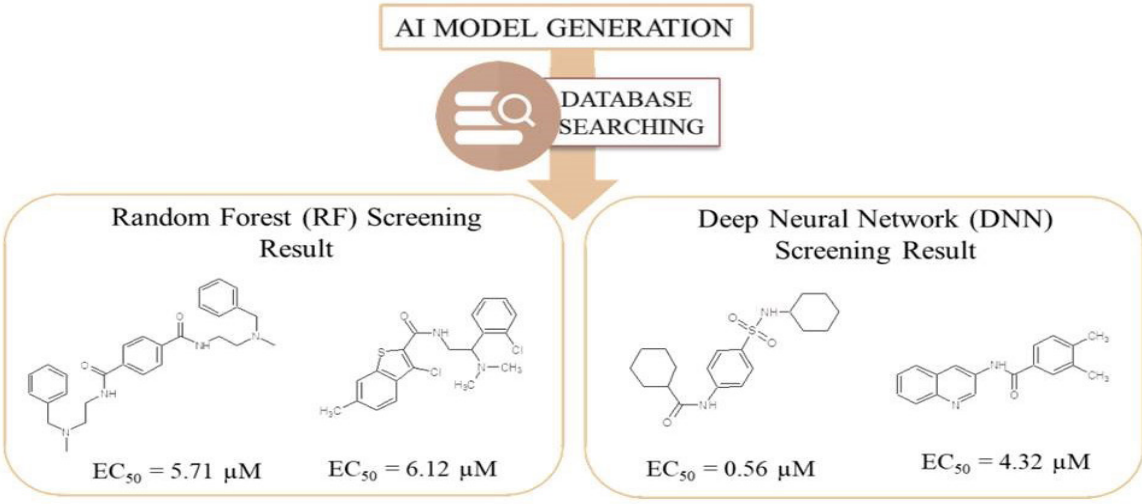

Figure 5. Tsou LK and coworkers' studies compounds (Tsou *et al.*, 2020)

In a different research, Koksal and Tugcu used QSARINS software to create QSAR models that predicted the analgesic and anti-inflammatory properties of several 2-benzoxazolinone derivatives. These models work based on the rule that the drug candidates' hydrophobicity, halogen count, and molecular structure shape are important indicators of their analgesic and anti-inflammatory properties. Seventy-seven novel compounds were introduced as possible analgesic and anti-inflammatory medications based on the previously investigated compounds and the models that were built. As a result, the majority of the recently developed compounds showed encouraging analgesic and anti-inflammatory effects (Tugcu and Koksal, 2018).

Table 3. AI-Based computational tools for hit-tolerated optimization

Tools	Description	Reference
DeepVS	Firstly, used deep learning to improve the performance of SBVS and used the DCNNs model. This method takes the result of molecular docking as the input of DCNN, and can automatically learn and extract relevant features.	Pereira *et al.*, 2016
BindScope	CNN-based protein-ligand docking and binding predictor	Skalic *et al.,* 2019
DeepConv-DTI	Provides prediction of drug-target interactions via deep learning with convolution on protein sequences.	Lee *et al.*, 2019
OntoQSAR	A machine learning model that interprets chemical and biological data in quantitative structure-activity relationship studies.	Angelo *et al.*, 2020
QSARINS	A software for the development and validation of multiple linear regression Quantitative Structure-Activity Relationship (QSAR) models by Ordinary Least Squares method and Genetic Algorithm for variable selection.	Gramatica *et al.*, 2013

SBVS: structure-based drug design,

DCNN: deep convolutional neural networks,

CNN: convolutional neural networks

Lead Optimization

Prediction of pharmacokinetic properties is not only crucial for decreasing the risks of late-stage drug development but also aids researchers in optimizing screening by prioritizing the testing of the most promising compounds. Absorption, distribution, metabolism, elimination, and toxicity (ADMET) prediction is an efficient method during hit-to-lead and lead optimization steps. It is documented that success rate and production efficiency in the drug development process mainly depend on the early estimation and optimization of the ADMET properties of the lead compounds. However, relying on *in vivo* experiments for estimating the ADMET of a compound comes with high costs, prolonged time requirements, and the need for substantial material and animal resources (Caldwell *et al*., 2009). The utilization of AI-assisted ADMET prediction has proven to be a cost-effective strategy, reducing drug development costs by up to 50%, making it a popular method in early drug discovery (Wang *et al*., 2019; Tan *et al.,* 2010). In addition, the success of AI-assisted ADMET prediction has been notably enhanced with the availability of high-quality data and more accurate statistical analysis methods. There are many approaches to ADMET modeling in drug discovery. Among ML methods, k-NN, SVM, RF and ANNs are used in the ADMET property investigation (Obrezanova *et al*., 2007, Kortagere *et al*., 2008, Cao *et al*., 2010, Cao *et al*., 2012, Klon *et al*., 2006, Lombardo *et al*., 2006, Wang *et al*., 2016). For example, DeepTox is a useful ML method that not only identifies static and dynamic properties within the chemical descriptors of the compounds but also predicts the toxicity of a molecule based on predefined 2500 toxicophore features (Mayr *et al*., 2016). A variety of AI tools used in ADMET prediction are detailed in Table 4.

Computer-aided AI methods are used for early and accurate prediction of adverse reaction of the drug compounds. For example, using in silico methods, Oner et al. analyzed the anticancer characteristics of Tetrahydrocannabinol (THC), Tetrahydrocannabivarin (THCV), and Cannabidiol (CBD). The ADMET properties of these compounds was evaluated using Protox-II. According to result, CBD has the lowest risk for immunotoxicity, carcinogenicity, and hepatotoxicity. On the other hand, the possibility of being inert in terms of mutagenicity and cytotoxicity is highest. Furthermore, CBD has the highest potential for preventing lung cancer (Gallerdo *et al*., 2024).

Günes and colleagues developed a model to predict 329 known antidepressant medications (ADRs) of which 27 were approved. They then looked at three ML algorithms (SVM, k-NN, and multilayer perception) to see which is more suitable for this task. To predict ADRs with AI model, they combined the chemical structures and biological properties (target protein, enzymes, and transporters) of compounds with the known ADRs of those drugs. The model they created using MLP with BestFirst and CfsSubsetEval, based on the chemical features of the approved antidepressants, correctly predicted the ADRs associated with the withdrawal of indalpine, zimelidine, pheniprazine, amineptine, and medifoxamine. The outcomes demonstrated the approach's external validity by correctly predicting a respectable number of previously identified ADRs from the literature (Günes *et al*., 2021).

Table 4. AI-Based computational tools for lead optimization

Tools	Description (available website)	Reference
ProTox-II	A webserver for the prediction of toxicity of chemicals. (http://tox.charite.de/protox_II)	Banerjee *et al*., 2018
ADMETlab	A platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. (http://admet.scbdd.com)	Dong *et al*., 2018
COSMOfrag	Provides its broad applicability for the accurate prediction of thermodynamic, environmental, or physiological properties.	Hornig and Klamt, 2005

ADMET: absorption, distribution, metabolism, excretion, and toxicity

The integration of AI in four essential stages, namely target identification, hit discovery, hit-to-lead optimization, and lead optimization, significantly reduces time and costs compared to traditional methods. Notably, some drugs designed through AI-driven approaches have received approval from the FDA for clinical trials (Liu *et al.*, 2019, Mak *et al.,* 2022, Soni *et al.*, 2022, Pun *et al.*, 2023). Since 2023, four prominent drugs that were FDA-approved for clinical trials are listed in Table 5.

Table 5. FDA-approved drugs for clinical trials

Drug	Clinical phase	Years when began the clinical trials	Therapeutical target	Pharmacological activity	Organization
DSP-1181	Phase 1	2020	5-HT1A receptor agonist	Obsessive-compulsive disorder	Exscientia and Sumitomo Dainippon Pharma
EXS21546	Phase 1	2020	Adenosine A2a receptor antagonist	Autoimmune oncology treatment	Exscientia and Evotec
DSP-0038	Phase 1	2021	Dual target on 5-HT1A receptor and 5HT2A receptor antagonist	Alzheimer's diseases	Exscientia and Sumitoto Dainippan Pharma
INSO18-055	Phase 2	2023	Anti-fibrotic small molecule inhibitor	Idiopathic Pulmonary Fibrosis	In silico Medicine

AI-Assisted Synthesis Planning

In drug discovery, the design of drug molecules must align with synthesizability to advance through the optimization process and yield compounds with improved properties. For this reason, organic synthesis is an essential part of drug discovery. Choosing the most suitable synthesis plan provides many advantages in terms of cost and time. Therefore, numerous computational approaches have been built to promote synthesis planning. There are three main tasks for AI-assisted synthesis: predicting retrosynthetic strategy, forecasting reaction conditions, and predicting side products of the selected reaction (Struble *et al*., 2020, Segler *et al*., 2018) (Figure 6).

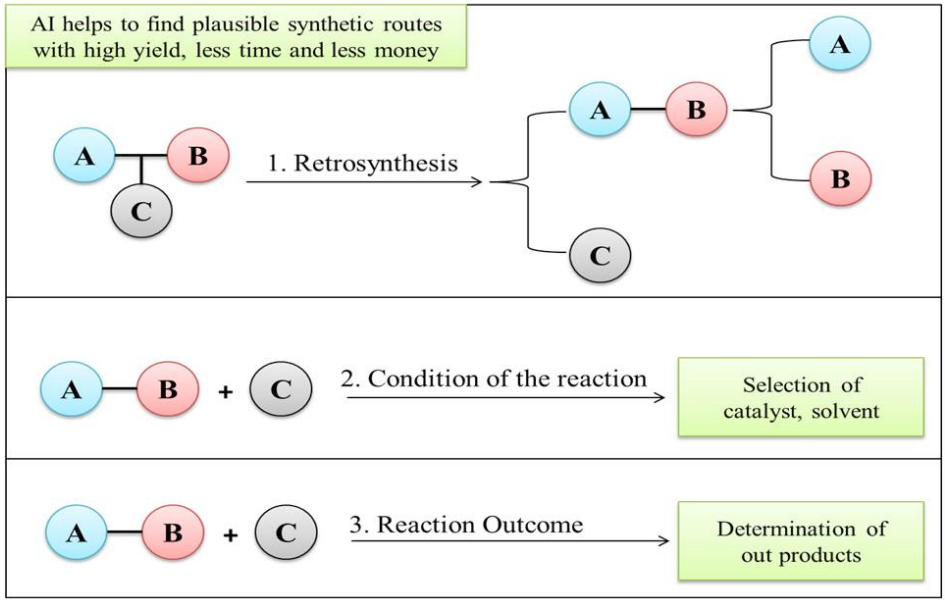

Figure 6. AI-assisted synthesis prediction

Retrosynthetic analysis is defined as transforming a target molecule into intermediates or precursors regardless of the reactivity reagents. Retrosynthesis pathway predictions cover the sequential cutting of the target compounds at different positions. Monte Carlo tree search (MCTS) is a method that was used in retrosynthesis prediction to perform branch decisions (Browne *et al.*, 2012). In recent years, various ML-based techniques have been introduced for retrosynthetic reaction prediction. For example, Liu *et al.* used a sequence-to-sequence-based model for retrosynthetic reaction prediction (Liu *et al.*, 2017). In 2017, Segler *et al.* utilized the first deep learning to find plausible synthetic routes with high yield and less time. Instead of manually encoding, a database of known reactions is provided to convert into reaction templates, considering the core of the reaction and the nearest neighbor atoms. The prediction of templates related to the targeted product in retrosynthetic analysis has been tackled using ANN. This approach enables the direct learning of retrosynthetic strategies through data. Furthermore, ANNs can select an efficient tree-search for a logical pathway among numerous reaction templates by filtering out results from implausible chemical reactions (Segler *et al.*, 2017). The work by Segler *et al.* demonstrated the feasibility of utilizing data-driven approaches, and this methodology has since been further enhanced with the availability of multiple open-sources as well as commercial tools (Table 6).

Table 6. AI-Based computational tools for synthesis planning

Tools	Description (available website)	Reference
Chemical.AI	A professional website to predict retrosynthesis routes.	https://chemical.ai
AiZynthFinder	A fast, robust, and flexible open-source software for retrosynthetic planning. (http://www.github.com/MolecularAI/aizynthfinder)	Genheden *et al.*, 2020
SciFinder	Make a whole retrosynthetic analysis powered by the renowned CAS collection of reactions, reducing the synthetic planning time.	Gabrielson., 2018
Reaxys	Provides prediction retrosynthesis combines high-quality reaction data with AI technology.	Goodman., 2009

According to CAS data, the SciFinder retrosynthesis planner builds pathways to desired compounds using experimental and predicted reaction steps from 121 million reactions in the CAS collection, which is compiled over 110 years of chemistry research. As an example, using SciFinder, the retrosynthesis of compound **A** is illustrated in Figure 7 (https://www.cas.org/resources/press-releases/scifinder-n-predictive-retrosynthesis).

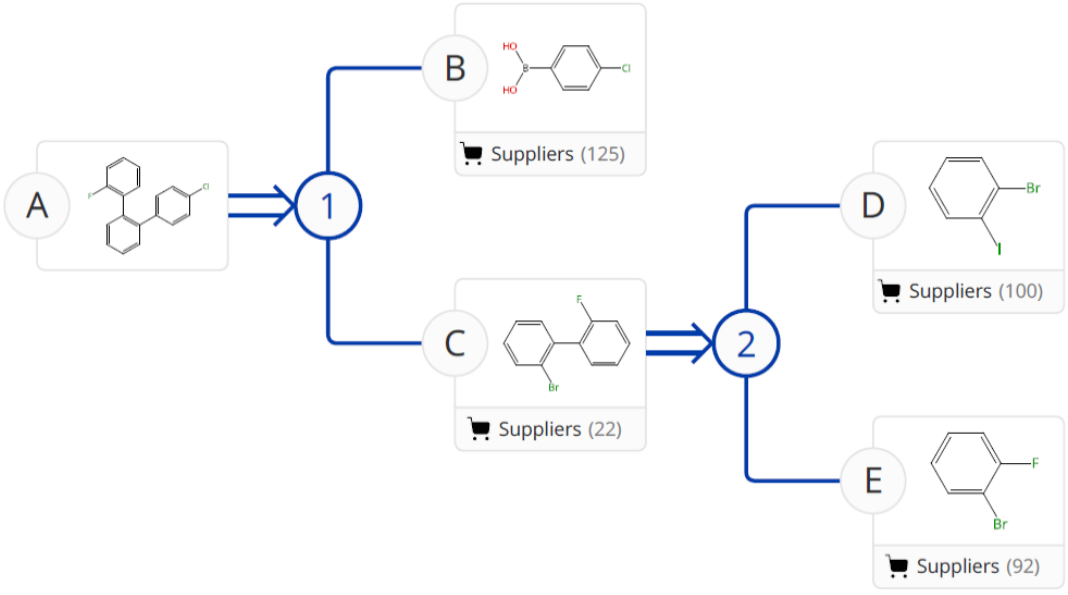

Figure 7. An example of a retrosynthesis tool of SciFinder produced at *https://scifinder-n.cas.org*

AI techniques are also effective in the prediction of the products and yields of organic reactions based on the molecular properties of the reactants. Recently, several studies on AI algorithms to generalize yield prediction have been documented (Hessler and Baringhaus, 2018, Struble *et al.*, 2020). Recurrent neural networks (RNNs) can be used to form *de novo* chemical discovery by using simplified molecular input line entry systems (SMILES) string representations of the structure. In chemical synthesis, starting materials and resulting compounds are encoded by SMILES strings and linked in an encoder-decoder architecture. The overall performance of this technology has demonstrated comparability to rule-based expert systems, although significant variations have been observed across different reaction classes (Savage *et al.*, 2017). An alternative approach involves the use of recommender systems to identify reactants yielding a desired product in combination with a chemical reaction graph. Differently, the recommender systems have been employed for identifying reactants yielding a targeted product with a chemical reaction graph. The utilization of deep neural networks combined with a Monte Carlo tree search provides an excellent performance for retrosynthetic prediction (Segler *et al.*, 2018).

Challenges of Using AI in Healthcare

Despite many advantages of bringing AI to drug discovery, it also has several challenges that must be considered (Blanco-Gonzalez *et al.*, 2023, Vamathevan *et al.*, 2019). The first challenge is suitable data availability. ML and DL approaches require a large volume of data for analyzing different tasks (Tsuji *et al.*, 2021). However, the accessible data can be restricted, or data can be low quality or inconsistent, leading to low accuracy and reliability of the results (Gomez *et al.*, 2018). Another difficulty arises from the fact that the data obtained in healthcare services are unclear,

noisy, and incomplete. This prevents AI from having clean and structured data and makes it difficult to apply AI in drug development (Manne and Kantheti, 2021). Finally, privacy and confidentiality are the most controversial topics of using AI in the drug discovery process. Providing personal information to any database and reusing this information without permission causes ethical problems (Aung *et al.,* 2021). To overcome this problem, measures protecting the confidentiality of patient information are necessary to be addressed within certain policies.

CONCLUSION

In summary, AI has attracted a lot of attention recently and has been successfully integrated into many steps of drug discovery. AI makes the drug discovery process shorter, cheaper, more advanced, and more reliable when compared to traditional drug discovery methods. In addition to assistance in quick and seamless identification of the hit compound, AI also contributes to the prediction of the potentially active drug candidate, understanding of drug-target interactions, ADMET properties, as well as suggestions of synthesis routes of active molecules. AI can also contribute to establishing the safety and efficacy of the product in clinical trials, as well as ensuring proper positioning and costing in the market through comprehensive market analysis and prediction. However, there are a limited number of AI-designed drugs that recived approval from FDA for clinical trials. Although specific challenges remain about the implementation of this technology, AI has gained a role as an invaluable tool in the pharmaceutical industry. All in all, AI serves as a golden key that has the potential to save lives by addressing critical segments of the drug discovery process, offering hope for numerous diseases that currently lack effective treatments or preventive measures.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION STATEMENT

Meric KOKSAL conceived of the presented idea. Meric KOKSAL encouraged Ayça DEDEOĞLU ERDOGAN, Armanç MAT and Enise Ece GURDAL to investigate and supervised the findings of this work. Conceptualization, K.M.; formal analysis, G.E.E.; writing-original draft preparation, D.E.A. and M.A ; writing-review and editing, K.M. and G.E.E.; visualization, K.M; supervision K.M. and G.E.E.; All authors have read and agreed to the published version of the manuscript.

REFERENCES

Abdel-Basset, M., Hawash, H., Elhoseny, M., Chakrabortty, R. K., & Ryan, M. (2020). DeepH-DTA: deep learning for predicting drug-target interactions: a case study of COVID-19 drug repurposing. *Ieee Access, 8,* 170433-170451.

Ain, Q. U., Aleksandrova, A., Roessler, F. D., & Ballester, P. J. (2015). Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 5(6),* 405-424. doi: 10.1002/wcms.1225.

Angelo, R. M., Io, A. K., Almeida, M. P., Silveira, R. G., Oliveira, P. R., Alcazar, J. J. OntoQSAR: An ontology for interpreting chemical and biological data in quantitative structure-activity relationship studies. *International Computer Science Conference, 203-206*, 2020, San Diego, CA, USA

Aoyama, T., Suzuki, Y., & Ichikawa, H. (1990). Neural networks applied to pharmaceutical problems. III. Neural networks applied to quantitative structure-activity relationship (QSAR) analysis. *Journal of Medicinal Chemistry, 33(9),* 2583-2590. doi: 10.1021/jm00171a037

Aung, Y. Y., Wong, D. C., & Ting, D. S. (2021). The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *British Medical Bulletin, 139(1),* 4-15. doi: 10.1093/bmb/ldab016.

Alamshany, Z. M., Khattab, R. R., Hassan, N. A., El-Sayed, A. A., Tantawy, M. A., Mostafa, A., & Hassan, A. A. (2023). Synthesis and molecular docking study of novel pyrimidine derivatives against COVID-19. *Molecules, 28(*2*),* 739. doi: 10.3390/molecules28020739.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., ... & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science, 373(*6557*),* 871-876. doi: 10.1126/science.abj8754

Ballester, P. J. (2019). Selecting machine-learning scoring functions for structure-based virtual screening. *Drug Discovery Today: Technologies, 32,* 81-87. doi: 10.1016/j.ddtec.2020.09.001

Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research, 46(*1*),* 257-263. doi: 10.1093/nar/gky318

Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., Conde-Torres, D., Antelo-Riveiro, P., Pineiro, A., & Garcia-Fandino, R. (2023). The role of ai in drug discovery: challenges, opportunities, and strategies. *Pharmaceuticals, 16(*6*),* 891.doi: 10.3390/ph16060891

Blundell, T. L. (1996). Structure-based drug design. *Nature, 384(*6604*),* 23. doi: 10.1038/384023a0

Bernard, S., Adam, S., & Heutte, L. (2012). Dynamic random forests. *Pattern Recognition Letters, 33(*12*),* 1580-1586. doi: 10.1016/j.patrec.2012.04.003

Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., ... & Colton, S. (2012). A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games, 4(*1*),* 1-43. doi: 10.1109/TCIAIG.2012.2186810

Caldwell, G. W., Yan, Z., Tang, W., Dasgupta, M., & Hasting, B. (2009). ADME optimization and toxicity assessment in early-and late-phase drug discovery. *Current Topics in Medicinal Chemistry, 9(*11*),* 965-980. doi: 10.2174/156802609789630929

Cano, G., Garcia-Rodriguez, J., Garcia-Garcia, A., Perez-Sanchez, H., Benediktsson, J. A., Thapa, A., & Barr, A. (2017). Automatic selection of molecular descriptors using random forest: Application to drug discovery. *Expert Systems with Applications, 72,* 151-159. doi: 10.1016/j.eswa.2016.12.008

Cao, D. S., Xu, Q. S., Liang, Y. Z., Chen, X., & Li, H. D. (2010). Prediction of aqueous solubility of druglike organic compounds using partial least squares, back-propagation network and support vector machine. *Journal of Chemometrics, 24(*9*),* 584-595.doi: 10.1002/cem.1321

Cao, D. S., Zhao, J. C., Yang, Y. N., Zhao, C. X., Yan, J., Liu, S., ... & Liang, Y. Z. (2012). In silico toxicity prediction by support vector machine and SMILES representation-based string kernel. *SAR and QSAR in Environmental Research, 23(*1-2*),* 141-153. doi: 10.1080/1062936X.2011.645874.

Cao, T., Wu, X., & Hu, X.. ProteinNET: A Protein Interaction Network Integration System. In Modern Advances in Intelligent Systems and Tool, 71-76, 2012, Berlin, Heidelberg.

Chan, H. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends in Pharmacological Sciences*, *40*(8), doi: 592-604. 10.1016/j.tips.2019.06.004

Cheng, F., & Sutariya, V. (2012). Applications of artificial neural network modeling in drug discovery. *Clinical and Experimental Pharmacology, 2(*3), 1-2. doi: 10.4172/2161-1459.1000e113

Coley, C. W., Eyke, N. S., & Jensen, K. F. (2020). Autonomous discovery in the chemical sciences part I: Progress. *Angewandte Chemie International Edition, 59(*51*),* 22858-22893. doi: 10.1002/anie.201909987

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20,* 273-297. Retrieved from https://link.springer.com/article/10.1007/BF00994018.

Dobchev, D., & Karelson, M. (2016). Have artificial neural networks met expectations in drug discovery as implemented in QSAR framework. *Expert Opinion on Drug Discovery, 11(7),* 627-639. doi: 10.1080/17460441.2016.1186876

Dobchev, D., G Pillai, G., & Karelson, M. (2014). In silico machine learning methods in drug development. *Current Topics in Medicinal Chemistry, 14(*16*),* 1913-1922. doi: 10.2174/1568026614666140929124203

Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D., ... & Cao, D. S. (2018). ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *Journal of Cheminformatics, 10*, 1-11. doi: 10.1186/s13321-018-0283-x

Doniger, S., Hofmann, T., & Yeh, J. (2002). Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms. *Journal of Computational Biology, 9(6),* 849-864. doi: 10.1089/10665270260518317.

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. doi: 10.2174/0929867043456287.

Dudek, A. Z., Arodz, T., & Gálvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Combinatorial Chemistry & High Throughput Screening, 9(*3*),* 213-228. doi: 10.2174/138620706776055539.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? *Springer International Publishing. 3-11,* 2015, Berlin, Germany.

Fahimian, G., Zahiri, J., Arab, S. S., & Sajedi, R. H. (2020). RepCOOL: computational drug repositioning via integrating heterogeneous biological networks. *Journal of Translational Medicine, 18(*1*),* 1-10. Retrieved from https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02541-3

Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature, 557(*7706*),* 55-55. Retrieved from http://www.nature.com/nature/index.html

Fuloria, N. K., Fuloria, S., & Vakiloddin, S. (2013). Phase zero trials: a novel approach in drug development process. *Renal Failure, 35(*7*),* 1044-1053. doi: 10.3109/0886022X.2013.810543

Gabrielson, S. W. (2018). *SciFinder. Journal of the Medical Library Association: JMLA, 106(*4*),* 588. doi: 10.5195/jmla.2018.515

Gallardo, A. A., Gutierrez, M. R., Gomez, L. A. J., Dumbrique, H. E. L., Liquido, M. I. H., Margaret, M., ... & Labrador, A. M. (2024). A Comparative Analysis on the Potential Anticancer Properties of Tetrahydrocannabinol, Cannabidiol, and Tetrahydrocannabivarin Compounds Through In Silico Approach. *Asian Pacific Journal of Cancer Prevention, 25(*3*),* 839-856. doi: 10.31557/APJCP.2024.25.3.839.

Genheden, S., Thakkar, A., Chadimová, V., Reymond, J. L., Engkvist, O., & Bjerrum, E. (2020). AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of Cheminformatics, 12(*1*),* 70. doi: 10.1186/s13321-020-00472-1

Ghasemi, F., Mehridehnavi, A., Pérez-Garrido, A., & Pérez-Sánchez, H. (2018). Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. *Drug Discovery Today, 23(*10*),* 1784-1790. doi: 10.1016/j.drudis.2018.06.016

Goodman, J. (2009). Computer Software Review, *Journal of Chemical Information and Modeling, 49,* 2897-2898. doi: 10.1021/ci900437n

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., ... & Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science, 4(*2*),* 268-276. doi: 10.1021/acscentsci.7b00572

Güneş, S. S., Yeşil, Ç., Gurdal, E. E., Korkmaz, E. E., Yarım, M., Aydın, A., & Sipahi, H. (2021). Primum non nocere: In silico prediction of adverse drug reactions of antidepressant drugs. *Computational Toxicology, 18,* 100165. doi: 10.1016/j.comtox.2021.100165

Gramatica, P., Chirico, N., Papa, E., Cassani, S., & Kovarich, S. (2013). QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *Journal of Computational Chemistry, 34(*24*),* 2121-2132. doi: 10.1002/jcc.23361

Heikamp, K., & Bajorath, J. (2014). Support vector machines for drug discovery. *Expert Opinion on Drug Discovery, 9(*1*),* 93-104. doi: 10.1517/17460441.2014.866943

Hessler, G., & Baringhaus, K. H. (2018). Artificial intelligence in drug design. *Molecules, 23(*10*),* 2520. doi: 10.3390/molecules23102520

Hornig, M., & Klamt, A. (2005). COSMO f rag: a novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. *Journal of Chemical Information and Modeling, 45(*5*),* 1169-1177. doi: 10.1021/ci0501948

Huang, K., Fu, T., Xiao, C., Glass, L., & Sun, J. (2020). Deep purpose: a deep learning based drug repurposing toolkit. *ArXiv Preprint ArXiv*:2004.08919. doi: 10.48550/arXiv.2004.08919

Jiang, L., Xu, M., Liu, T., Qiao, M., & Wang, Z. Deepvs: A deep learning based video saliency prediction approach. *In Proceedings of the European Conference on Computer Vision (eccv),* 602-617, 2018, Munich, Germany.

Jiménez, J., Skalic, M., Martinez-Rosell, G., & De Fabritiis, G. (2018). K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of Chemical Information and Modeling, 58(*2*),* 287-296. doi: 10.1021/acs.jcim.7b00650

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596(*7873*),* 583-589. Retrieved from https://www.nature.com/articles/s41586-021-03819-2

Klon, A. E., Lowrie, J. F., & Diller, D. J. (2006). Improved naive Bayesian modeling of numerical data for absorption, distribution, metabolism and excretion (ADME) property prediction. *Journal of Chemical Information and Modeling, 46(*5*),* 1945-1956. doi: 10.1021/ci0601315

Kortagere, S., Chekmarev, D., Welsh, W. J., & Ekins, S. (2008). New predictive models for blood–brain barrier permeability of drug-like molecules. *Pharmaceutical Research, 25,* 1836-1845. doi: 10.1007/s11095-008-9584-5

Lavecchia, A. (2015). Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today, 20(*3*),* 318-331. doi: 10.1016/j.drudis.2014.10.012

Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current Medicinal Chemistry, 20(*23*),* 2839-2860. doi: 10.2174/09298673113209990001

Lee, I., Keum, J., & Nam, H. (2019). DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. PLoS *Computational Biology, 15(6),* e1007129. doi: 10.1371/journal.pcbi.1007129

Lewis, T., & Writer, S. (2014). A brief history of artificial intelligence. *Live Science, 61(*4*),* 000812561986492 doi: 10.1177/0008125619864925

Li, H., Hou, J., Adhikari, B., Lyu, Q., & Cheng, J. (2017). Deep learning methods for protein torsion angle prediction. *BMC Bioinformatics, 18(*1*),* 1-13. Retrieved from https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1834-2

Li, H., Sze, K. H., Lu, G., & Ballester, P. J. (2020). Machine-learning scoring functions for structure-based drug lead optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 10(*5*),* e1465. doi: 10.1002/wcms.1465

Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1993). DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial intelligence, 61(*2*),* 209-261. doi: 10.1016/0004-3702(93)90068-M

Liu, B., He, H., Luo, H., Zhang, T., & Jiang, J. (2019). Artificial intelligence and big data facilitated targeted drug discovery. *Stroke and Vascular Neurology, 4(*4*),* 206-213. doi: 10.1136/svn-2019-000290.

Liu, B., Ramsundar, B., Kawthekar, P., Shi, J., Gomes, J., Luu Nguyen, Q., ... & Pande, V. (2017). Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science, 3(*10*),* 1103-1113. doi: 10.1021/acscentsci.7b00303

Lombardo, F., Obach, R. S., DiCapua, F. M., Bakken, G. A., Lu, J., Potter, D. M., ... & Zhang, Y. (2006). A hybrid mixture discriminant analysis–random forest computational model for the prediction of volume of distribution of drugs in human. *Journal of Medicinal Chemistry, 49(*7*),* 2262-2267. doi: 10.1021/jm050200r

Madhukar, N. S., Khade, P. K., Huang, L., Gayvert, K., Galletti, G., Stogniew, M., ... & Elemento, O. (2017). A new big-data paradigm for target identification and drug discovery. *Biorxiv,* 134973. doi: 10.1101/134973

Mak, K. K., Balijepalli, M. K., & Pichika, M. R. (2022). Success stories of AI in drug discovery-where do things stand?. *Expert Opinion on Drug Discovery, 17(*1*),* 79-92. doi: 10.1080/17460441.2022.1985108

Manne, R., & Kantheti, S. C. (2021). Application of artificial intelligence in healthcare: chances and challenges. *Current Journal of Applied Science and Technology, 40(*6*),* 78-89. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4393347

Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2016). DeepTox: toxicity prediction using deep learning. *Frontiers in Environmental Science, 3,* 80. Retrieved from https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2015.00080/full

Melville, J. L., Burke, E. K., & Hirst, J. D. (2009). Machine learning in virtual screening. *Combinatorial Chemistry & High Throughput Screening, 12(*4*),* 332-343. doi: 10.2174/138620709788167980

Ning, X., & Karypis, G. (2011). In silico structure-activity-relationship (SAR) models from machine learning: a review. *Drug Development Research, 72(*2*),* 138-146. doi: 10.1002/ddr.20410

Obrezanova, O., Csányi, G., Gola, J. M., & Segall, M. D. (2007). Gaussian processes: a method for automatic QSAR modeling of ADME properties. *Journal of Chemical Information and Modeling, 47(*5*),* 1847-1857. doi: 10.1021/ci7000633

Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS (2002). Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins, 46,* 34-40. doi: 10.1002/prot.10028

Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical Reviews, 9,* 91-102. doi: 10.1007/s12551-016-0247-1

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today, 26(*1*),* 80. doi: 10.1016/j.drudis.2020.10.010

Pereira, J. C., Caffarena, E. R., & Dos Santos, C. N. (2016). Boosting docking-based virtual screening with deep learning. *Journal of Chemical Information and Modeling, 56(*12*),* 2495-2506. doi: 10.1021/acs.jcim.6b00355

Pun, F. W., Ozerov, I. V., & Zhavoronkov, A. (2023). AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences, 44(9),* 561-572. doi: 10.1016/j.tips.2023.06.010

Rao, V. S., & Srinivas, K. (2011). Modern drug discovery process: An in silico approach. *Journal of Bioinformatics and Sequence Analysis, 2(*5*),* 89-94. Retrieved from http://www.academicjournals.org/JBSA

Repasky, M. P., Murphy, R. B., Banks, J. L., Greenwood, J. R., Tubert-Brohman, I., Bhat, S., & Friesner, R. A. (2012). Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *Journal of Computer-aided Molecular Design, 26,* 787-799. doi: 10.1007/s10822-012-9575-9.

Reynolds CH, Merz KM, Ringe D, eds. (2010). Drug Design: Structure- and Ligand-Based Approaches (1 ed.). *Cambridge, UK: Cambridge University Press.* ISBN 978-0521887236.

Savage, J., Kishimoto, A., Buesser, B., Diaz-Aviles, E., & Alzate, C. Chemical reactant recommendation using a network of organic chemistry. I*n Proceedings of the Eleventh ACM Conference on Recommender Systems,* 210-214, 2017, Como, Italy.

Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery, 4(*8*),* 649-663. Retrieved from https://www.nature.com/articles/nrd1799

Segler, M. H., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature, 555(*7698*),* 604-610. Retrieved from https://www.nature.com/articles/nature25978

Segler, M., Preuß, M., & Waller, M. P. (2017). Towards" alphachem": Chemical synthesis planning with tree search and deep neural network policies. A*rXiv Preprint ArXiv:*1702.00020. doi: 10.48550/arXiv.1702.00020

Shen, C., Ding, J., Wang, Z., Cao, D., Ding, X., & Hou, T. (2020). From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 10(*1*),* e1429. doi: 10.1002/wcms.1429

Skalic, M., Martínez-Rosell, G., Jiménez, J., & De Fabritiis, G. (2019). PlayMolecule BindScope: large scale CNN-based virtual screening on the web. *Bioinformatics, 35(7),* doi: 1237-1238. 10.1093/bioinformatics/bty758

Soni, K., & Hasija, Y. Artificial Intelligence Assisted Drug Research and Development. In 2022 IEEE Delhi Section Conference (DELCON), pp. 1-10, 2022, New Delhi, India.

Spencer, M., Eickholt, J., & Cheng, J. (2014). A deep learning network approach to ab initio protein secondary structure prediction. IEEE/ACM Transactions on Computational Biology and *Bioinformatics, 12(*1*),* 103-112. doi: 10.1109/TCBB.2014.2343960

Struble, T. J., Alvarez, J. C., Brown, S. P., Chytil, M., Cisar, J., DesJarlais, R. L., ... & Jensen, K. F. (2020). Current and future roles of artificial intelligence in medicinal chemistry synthesis. *Journal of Medicinal Chemistry, 63(*16*),* 8667-8682. doi: 10.1021/acs.jmedchem.9b02120

Tan, J. J., Cong, X. J., Hu, L. M., Wang, C. X., Jia, L., & Liang, X. J. (2010). Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection. *Drug Discovery Today, 15(*5-6*),* 186-197. doi: 10.1016/j.drudis.2010.01.004

Tsou, L. K., Yeh, S. H., Ueng, S. H., Chang, C. P., Song, J. S., Wu, M. H., ... & Ke, Y. Y. (2020). Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery. *Scientific Reports, 10(*1*),* 16771. Retrieved from https://www.nature.com/articles/s41598-020-73681-1

Tsuji, S., Hase, T., Yachie-Kinoshita, A., Nishino, T., Ghosh, S., Kikuchi, M., ... & Tanaka, H. (2021). Artificial intelligence-based computational framework for drug-target prioritization and inference of novel repositionable drugs for Alzheimer's disease. *Alzheimer's Research & Therapy, 13(*1*),* 1-15. Retrieved from https://alzres.biomedcentral.com/articles/10.1186/s13195-021-00826-3

Tugcu, G., & Koksal, M. (2019). A QSAR Study for Analgesic and Anti-inflammatory Activities of 5-/6-Acyl-3-alkyl-2-Benzoxazolinone Derivatives. *Molecular Informatics, 38(*8-9*),* 1800090. doi: 10.1002/minf.201800090

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery, 18(*6*),* 463-477.

Wang, C., & Zhang, Y. (2017). Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. *Journal of Computational Chemistry, 38(*3*),* 169-177. doi: 10.1002/jcc.24667

Wang, L., Ding, J., Pan, L., Cao, D., Jiang, H., & Ding, X. (2019). Artificial intelligence facilitates drug design in the big data era. *Chemometrics and Intelligent Laboratory Systems, 194,* 103850. doi: 10.1016/j.chemolab.2019.103850

Wang, N. N., Dong, J., Deng, Y. H., Zhu, M. F., Wen, M., Yao, Z. J., ... & Cao, D. S. (2016). ADME properties evaluation in drug discovery: prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting. *Journal of chemical information and modeling, 56(*4*),* 763-773. doi: 10.1021/acs.jcim.5b00642

Wang, T., Qiao, Y., Ding, W., Mao, W., Zhou, Y., & Gong, H. (2019). Improved fragment sampling for ab initio protein structure prediction using deep neural networks. *Nature Machine Intelligence, 1(*8*),* 347-355. doi: 10.1038/s42256-019-0075-7

Zhong, F., Xing, J., Li, X., Liu, X., Fu, Z., Xiong, Z., ... & Jiang, H. (2018). Artificial intelligence in drug design. *Science China Life Sciences, 61,* 1191-1204. doi: 10.1007/s11427-018-9342-2

Zhu, W., Liu, X., Li, Q., Gao, F., Liu, T., Chen, X., ... & Zhavoronkov, A. (2023). Discovery of novel and selective SIK2 inhibitors by the application of AlphaFold structures and generative models. *Bioorganic & Medicinal Chemistry, 91*, 117414. doi: 10.1016/j.bmc.2023.117414