Nirma Univ J Pharm Sci; 2021, 8(1&2) 37-44

ISSN 2348 –4012

ARTICLE

CASE STUDIES ON APPLICATIONS OF COMPUTATIONAL TECHNIQUES IN DRUG DESIGN

Ekta Thakor[1,2], Bhumika Patel[1]*

[1] Department of Pharmaceutical Chemistry, Institute of Pharmacy, Nirma University, S. G. Highway, Ahmedabad, Gujarat, India

[2]St. John Institute of Pharmacy and Research, Palghar, Maharashtra, India

Corresponding author:

Bhumika Patel

Email: bhumipatel2212@gmail.com, bhumika.patel@nirmauni.ac.in

ABSTRACT

Computer-aided drug design (CADD) along with Artificial Intelligence (AI) based machine learning technologies have a powerful impact in the field of drug discovery as it can handle vast biological data which in turn reduces the cost and time of drug discovery and development process. Identifying hits through virtual screening and its further optimization for the development of lead molecule through ligand- or structure-based drug designing have played a vital role. Docking and molecular dynamics studies highlighted the binding of ligands with targeted proteins and their binding affinity. ADMET prediction studies prevent the failure of many drugs in clinical trials and thus prevent loss of time and money. The availability of open-source big data has facilitated the screening of vast libraries intending to come up with novel potent and target-specific drugs. Present review has given an overview of the technology used in drug discovery by highlighting some of the case studies.

Keywords: *Artificial Intelligence (AI), Machine Learning (ML), Ligand- based drug design (LBDD), and Structure-based drug design (SBDD)*

INTRODUCTION

Drug discovery is an interdisciplinary multiple-step process that is time-consuming and expensive [1]. It requires 15-20 years for a drug to reach the market from the concept; which includes phases like target identification and validation, lead identification and its synthesis, pre-clinical screening, clinical trials and regulatory approval [2,3]. Although a huge amount of money and time are involved in the drug discovery process, the success rate is as low as 15-20%. The major cause of the failure is the poor ADMET profile of the lead [4].

To accelerate and elevate the success rate in the drug discovery process, the computational approach is harnessed to a tremendous extent [5]. With the swift evolution of diverse computer hardware and software, various structure based and ligand-based drug design techniques or algorithms are being utilized nowadays in drug screening and design [6]. For the advanced *in silico* drug design, chemoinformatic is considered as a prime tool [7]. Bioactivity spectra-based algorithms, data mining, chemical structure similarity searching, panel docking etc. are some representative examples of cheminformatics tools [8]. Target identification which is the first step in rational drug design; can be studied through network pharmacology, a field assimilating varied information in protein-disease, protein-protein and drug-protein networks [9]. Network pharmacology is very helpful in study of inter-correlations across different biological pathways, genomics, proteomics, metabolomics, transcriptomics etc. that rely on computational data and its interpretation [10]. Other approaches are; protein ligand interaction fingerprints method, ligand-based interaction fingerprint method, MetaboAnalyst approach for pathway analysis etc. [11,12].

Present review summarizes the published case studies on the basis of type of application of computational techniques and artificial intelligence in the drug design as summarized in **Fig. 1**.

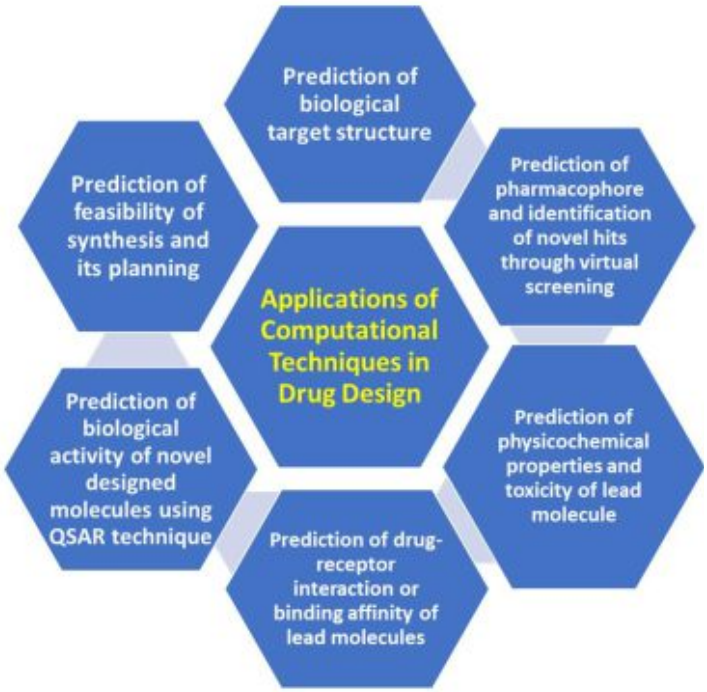

Fig. 1 Diverse Applications of Computational Techniques in Drug Design

APPLICATIONS OF COMPUTATIONAL TECHNIQUES IN DRUG DESIGN

Prediction of biological target structure

For structure-based drug design (SBDD), availability of 3D crystal structure biological macromolecular target is crucial. Through the advancements in the structure elucidating experimental techniques like X-ray, NMR, electron microscopy etc., 3D structure of proteins is resolved and then they are made available in public for the users [13]. However, in cases where 3D structure of the target protein is not known, computational techniques like homology modeling or ab initio modeling can be used to predict the 3D protein structure [14].

Homology modeling is a computational technique to predict protein structure based on its sequence identity and similarity with the template protein of known structure. For the reliable prediction, minimum 30% sequence identity should be possessed between the query sequence and template sequence [15].

Recently in 2024, Goswami, V. et al. [16] have reported the homology modeling of Porcupine protein which an enzyme involved in the Wnt signaling pathway. Various disorders like osteoporosis, cancer,

Alzheimer's disease etc. are associated with the dysregulated Wnt signaling. To design novel Porcupine inhibitors, there was a need of 3D structure of Porcupine. Goswami, V. et al. have predicted its 3D structure through homology modeling using an online server, I-TASSER and commercial software, Molsoft ICMPro. They compared and validated both the homology models through Ramachandran plot, Protein health tool of Molsoft ICM and other tools available on metaserver, SAVES v6.0. Molsoft model was found to be better with 84.6 % residues in most favored region and only 0.3 % residues in disallowed region in Ramachandran plot in comparison to 75.9% and 1.7% residues respectively in I-TASSER model. Finally, the predicted model using Molsoft ICMPro was considered for molecular docking of known porcupine inhibitors and future designing of novel Porcupine inhibitors. Later, authors have also compared the homology model binding site with that of the actual binding site of Porcupine structure deposited in the protein databank; PDB ID: 7URD and found high similarity between them. Thus, Computational technique, homology modeling was proved crucial in the designing of novel inhibitors.

Prediction of physicochemical properties and toxicity of lead molecule

Prediction of pharmacokinetic properties is an integral part of drug discovery that uses diverse AI-based tools which can analyze the behavior of drugs in the body. Few structural parameters being used in drug designing are molecular descriptors like SMILES strings, electron distribution of the molecule, energy calculation and bonds etc. These parameters are used in deep neural networking (DNN) for drug designing [17]. (Yukawa & Naven, 2020).

In 2019, Han et. al. [18] assessed the *in silico* ADMET profile of ceftazidime as well as its impurities A-I using Discover Studio 4.0 (DS4.0) software package and pkCSM ADMET descriptors algorithm protocol. Ceftazidime has been in clinical use since 1990s to treat various infections. There are many adverse effects related to its use which are not only related to the toxicity of API but also due to impurities in drugs. To identify the type of impurity and its structural features responsible for toxicity, prediction of ADMET properties of all the known impurities A-I have been carried out by Han et. al. Through structure toxicity relationship study, they could identify specific functional groups and stereochemistry of the drug, impurities and its metabolites responsible for various types of toxicity like neurotoxicity, genotoxicity etc. This study overall provided the base for experimental validation of predicting the toxicity of all other cephalosporins and their impurities, and for the quality control of these impurities.

Prediction of drug-receptor interaction or binding affinity of lead molecules/drugs

Molecular docking is an interesting in silico technique to determine the interaction of ligands with macromolecules free binding energy. It requires an input of 3D structure of an unbound macromolecule either directly obtained from protein data bank (PDB) or through homology modeling to predict the covalent/non-covalent binding of ligand molecules.

In 2021 Dassi et al. [19] reported drug repurposing for Leishmaniasis through hybrid approach. They blended representation learning, a deep learning approach with molecular docking to predict ligand-target interaction. Initially, an online deep learning-based tool, DeepPurpose, was used for docking based virtual screening of large dataset total 2,058,752 protein-ligand pairs which shortlisted 3400 pairs based on the affinity scores. These pairs were then processed for computing interface energies using molecular docking by Autodock Vina. The deep learning assisted hybrid approach proved to be 50 times faster than conventional docking based virtual screening without losing valuable drug candidates. The model predicted alpha-glutamicin, a non-steroidal anti-inflammatory drug, as a promising repurposed drug for the treatment of leishmaniasis.

Prediction of pharmacophore and identification of novel hits through virtual screening

According to the IUPAC [20], "A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supra molecular interactions with a specific biological target structure and to trigger (or to block) its biological response". Various pharmacophoric features are hydrogen bond donor (D), hydrogen bond acceptor (A), ring aromatic (RA), hydrophobic (H), Positive ionizable (PI), negative ionizable (NI) etc.

Babu et al. in 2022 [21] reported a study for identification of novel lead molecules as Janus kinase 1 (JAK1) inhibitors which can be used for the treatment of autoimmune diseases and cancer. They adopted hybrid approach and performed ligand-based pharmacophore modelling, virtual screening and molecular docking studies. Using 52 reported C-2 methyl/hydroxyethyl imidazopyrrolopyridines derivatives, ligand-based pharmacophore models were generated by Phase 4.3 module of a software, Schrodinger. The top 8 models were selected on the basis of validation techniques; Guner-Henry score and selectivity check for virtual screening of

various chemical libraries like Chemdiv, Asinex, Maybridge, Zinc, Enamine and Lifechemicals. ADMET study was performed on the identified hit and finally 2,856 hits which showed acceptable drug-like properties were selected for induced fit docking, MM-GBSA calculation and cross docking using GLIDE. Later, molecular dynamics (MD) simulation study was performed for the top 5 candidates using the GROMACS and DFT calculations using Gaussian. At last, best 2 hits, T5923531 and T5923555 were concluded as the best molecules for experimental validation using *in vitro* and *in vivo* techniques.

Prediction of biological activity of novel designed molecules using QSAR technique

Quantitative structure activity relationship (QSAR) is a ligand based drug design technique that establishes the relationship between different 2D/3D physico chemical descriptors of ligands with biological activity quantitatively using statistical methods and predict the activity of designed molecules before their synthesis [22,23]. Recently AI techniques are incorporated to develop SAR/QSAR predictive models which not only reduces the time but also allows the input selection to extract the features that affect the generated model. It improves the prediction accuracy and efficiency of QSAR [24].

Tsou L.K. et al. [25] in 2020 reported a comparative study between the deep neural networks (DNN) based QSAR model and PLS based QSAR model for their hit prediction efficiency thorough virtual screening. They proved DNN as well as random forest (RF) techniques superior. A database of 7130 molecules reported for their MDA-MB-231 inhibitory activities were collected from ChEMBL website and divided into training set (85%, 6069 compounds) and test set (15%, 1061 compounds) to develop the QSAR models. The results showed that DNN based method could achieve higher predicted r^2 value with only fewer compounds (Only 63) required in the training set than traditional QSAR method. Using this model, a potent (~ 500 nM) mu-opioid receptor agonist was identified as a hit from the in-house database of 165,000 compounds. This study proved the efficiency of AI based QSAR techniques in novel hit identification and prediction of their activity.

Prediction of feasibility of synthesis and its planning

The traditional path of drug discovery is organic synthesis but it is often accompanied by synthetic challenges which restrict chemical space available for drug designing [26]. Varied computational

approaches have been developed for doing systematic synthetic planning. Three aspects of the synthetic scheme are emphasized in computational studies, predicting the compound structure based on the starting chemical structure; predicting practical yield; and retrosynthetic planning [27]. In retrosynthetic planning, a knowledge-based system follows the rule from the reaction database. Forward synthesis prediction ranks the synthetic routes with a Monte Carlo tree search from retrosynthetic analysis using ML-based approach for an excellent performance. One of the open-source software is AiZynthFinder for retrosynthetic planning. In this software, Monte Carlo tree search system guided by an artificial neural network approach develops an algorithm that breaks down a molecule to purchasable chemical reagents and starting materials [28].

Thakkar A et al. [29] developed a basic retrosynthetic tool based on single neural network to investigate the role of the machine learning template prioritization method in the tree search algorithm. They trained USPTO dataset and developed a deep learning-based model and predicted the retrosynthesis route for a drug, Amenamevir. The route suggested by the model was compared with the literature routes. The model predicted the route of synthesis in just 4.26 seconds.

CONCLUSION:

With the case studies discussed above we can conclude that use of computational technology in drug discovery is a vital and appreciated tool that reduces the cost and time for research in the field of drug discovery. In the Big Data Era, hybridization of Artificial intelligence, machine learning, and deep learning approaches result in dealing with massive amounts of data and enable us to make quick decision. As we understand the application of AI for handling big data, we need to take care to avoid the access and utility of poor-quality data as input which hampers those final results of drug discovery related computational studies. The validation the hypothesis generated from computational approach is essential and inevitable steps. As the same time precaution of should be taken for distinguish false positive result from true positive results.

ACKNOWLEDGMENT:

Ekta Thakor and Bhumika Patel are thankful to Nirma University, Ahmedabad, Gujarat, India for providing resourceful support to carry out literature review for the present work.

REFERENCES:

[1] Berdigaliyev, N., Aljofan, M. (2020). An overview of drug discovery and

development. *Future Medicinal Chemistry*, *12*(10), 939–947. https://doi.org/10.4155/fmc-2019-0307

[2] Chua, H. M., Moshawih, S., Goh, H. P., Ming, L. C., Kifli, N. (2023). Insights into the computer-aided drug design and discovery based on anthraquinone scaffold for cancer treatment: A protocol for systematic review. *PLOS ONE*, *18*(9), e0290948. https://doi.org/10.1371/journal.pone.0290948

[3] Talevi, A. (2024). Computer-Aided Drug Discovery and Design: Recent Advances and Future Prospects. In M. Gore & U. B. Jagtap (Eds.), *Computational Drug Discovery and Design* (pp. 1–20). Springer US. https://doi.org/10.1007/978-1-0716-3441-7_1

[4] Vemula, D., Jayasurya, P., Sushmitha, V., Kumar, Y. N., Bhandari, V. (2023). CADD, AI and ML in drug discovery: A comprehensive review. *European Journal of Pharmaceutical Sciences*, *181*, 106324. https://doi.org/10.1016/j.ejps.2022.106324.

[5] Bajorath, J. (2022). Deep Machine Learning for Computer-Aided Drug Design. *Frontiers in Drug Discovery*, *2*, 829043. https://doi.org/10.3389/fddsv.2022.829043

[6] Selvaraj, C., Chandra, I., Singh, S. K. (2022). Artificial intelligence and machine learning approaches for drug design: Challenges and opportunities for the pharmaceutical industries. *Molecular Diversity*, *26*(3), 1893–1913. https://doi.org/10.1007/s11030-021-10326-z

[7] Chen, H., Kogej, T., Engkvist, O. (2018). Cheminformatics in Drug Discovery, an Industrial Perspective. *Molecular Informatics*, *37*(9–10), e1800041. https://doi.org/10.1002/minf.201800041.

[8] Arabi, A. A. (2021). Artificial intelligence in drug design: Algorithms, applications, challenges and ethics. *Future Drug Discovery*, *3*(2), 59. https://doi.org/10.4155/fdd-2020-0028

[9] Rasul, A., Riaz, A., Sarfraz, I., Khan, S. G., Hussain, G., Zara, R., Sadiqa, A., Bushra, G., Riaz, S., Iqbal, M. J., Hassan, M., Khorsandi, K. (2022). Target Identification Approaches in Drug Discovery. In M. T. Scotti & C. L. Bellera (Eds.), *Drug Target Selection and Validation*, 41–59. Springer International Publishing. https://doi.org/10.1007/978-3-030-95895-4_3

[10] Katsila, T., Spyroulias, G. A., Patrinos, G. P., Matsoukas, M-T. (2016).