Receptor.AI Tests an Innovative Technique of Chemical Data Augmentation for Discovering Anti-cancer Drugs

Feb. 4, 2022   by Alan Nafiiev

The BRD4 (Bromodomain-containing protein 4) protein is a transcriptional factor that plays an important role in cancer development. It has attracted a lot of attention in recent years as a promising drug target because its inhibition stops the proliferation of cancer cells.

BRD4 is a member of a family which contains several related proteins sharing the same structural motif called Bromo- and Extra-Terminal domain (BET). Although dozens of drugs that target bromodomains are proposed, the search for better inhibitors of BRD proteins in general and selective inhibitors of BRD4 in particular is still of great demand for cancer therapy.

Receptor.AI, a startup company focused on AI-based drug discovery, has chosen BRD4 as one of the targets to test the company's innovative technologies. The existence of a large number of known inhibitors for BRD proteins allowed it to utilize the ligand-based drug discovery route. In this approach all known ligands of the target protein, along with their activities and binding properties, are used as a training dataset for an Artificial Intelligence model. The model is then able to screen a large chemical library and detect the most promising molecules, which are called the hit compounds.
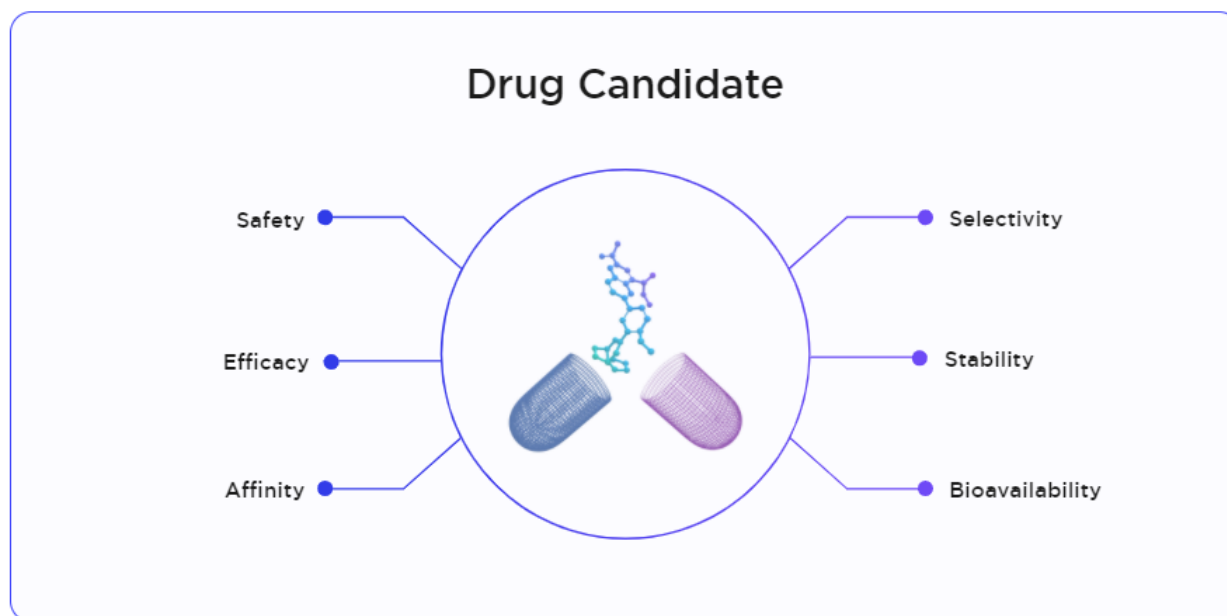
The ligand-based approach to AI-driven drug discovery is not new and is used by many companies worldwide with mixed results. However, Receptor.AI enhanced it with a proprietary technique called pharmacophore-based augmentation.

The quality of the ligand-based AI-models depend critically on the number of available ligands and their chemical diversity. Usually, the number of known ligands is small and the resulting model appears to be significantly limited and biased, which decreases its utility and accuracy.

An approach of Receptor.AI allows getting more training data from the same set of molecules by accounting not only for their chemical connectivity, but also for spatial structures and flexibility. It is based on pharmacophores — molecular descriptors, which encode the physical and chemical properties of chemical groups. The sequence of pharmacophores constitutes a unique fingerprint, which depends on the chemical identity and spatial pose of the molecule. For each ligand in the available training set a number of 3D conformers is generated. The conformation-sensitive pharmacophores are computed for each of them and encode them into the molecular representation suitable for AI training.

Such augmentation of the initial set of ligands dramatically increases the amount of data available for ML and improves the quality of the model, which is capable of recognizing the spatial patterns of pharmacophores. As a result, it works reliably even on the molecules that are chemically distant from everything presented in the training dataset — the feature is invaluable for screening large and diverse chemical databases.

This technological advancement allowed Receptor.AI to find novel inhibitors of BRD4 in less than two months including experimental validation. First, the AI model based on the pharmacophore-augmented dataset of BDR inhibitors was trained. Then the database of 3.2M high-quality drug-like molecules was screened using this model. 100 best-ranked hit compounds were passed to experimental validation to Receptor.AI's partners. 17 of them have shown affinity to BRD4 and 7 were binding to this protein selectively. These compounds are currently subject to in depth experimental validation followed by hit to lead and lead optimization stages.

- RECEPTOR.AI