AI-powered module for novel chemical entity generation

MEDVOLT

504, ZENNIA COMPLEX,
KOTHRUD, PUNE - 411038

www.medvolt.ai

Contents:

Introduction

Big picture of Novel Chemical Entity(NCE) module

Fragment-based de novo design

AI in de novo design (AIDD)

Filtering & Hit Identification

Hit Optimization

Outcome

Why do you want to use our NCE module?

Introduction

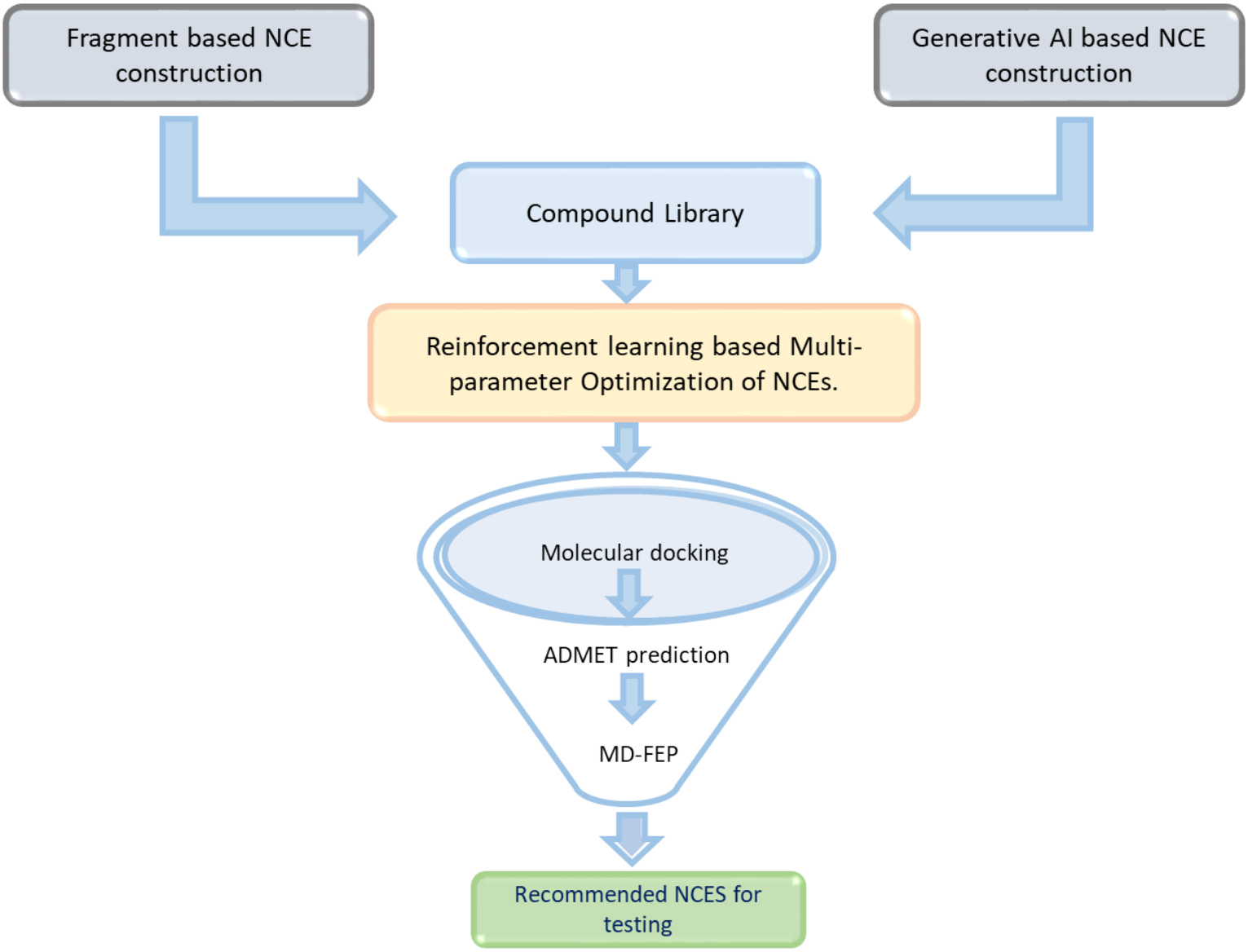
The term de novo design defines the utilization of computational methods to generate new compounds from scratch in a completely automated manner. The main advantage of de novo design includes the exploration of huge chemical space for developing more potent compounds with desired properties in lesser time and cost. Though the technique was considered difficult in the past, thanks to the introduction of artificial intelligence, the de novo drug design is getting popular now.

The overall expense of drug discovery has been cut down to nearly 70% with the introduction of artificial intelligence in the drug discovery pipeline. Also, the Inclusion of AI in the drug discovery pipeline accelerates the discovery process and literature shows that a biotechnology company was able to identify a novel drug candidate in just 8 months using its AI design platform [1].

Along with the conventional de novo design methodologies, **Medvolt's** NCE pipeline incorporates AI in the design platform. **Medvolt's** NCE module is highly customizable, our customer can predefine and set the value of molecular properties apriori, and our module guarantees the generation of the compound with a defined set of values. Along with the advanced approaches, our module can develop compounds with greater stability, selectivity, efficacy, synthesizability, and with the right amount of novelty.

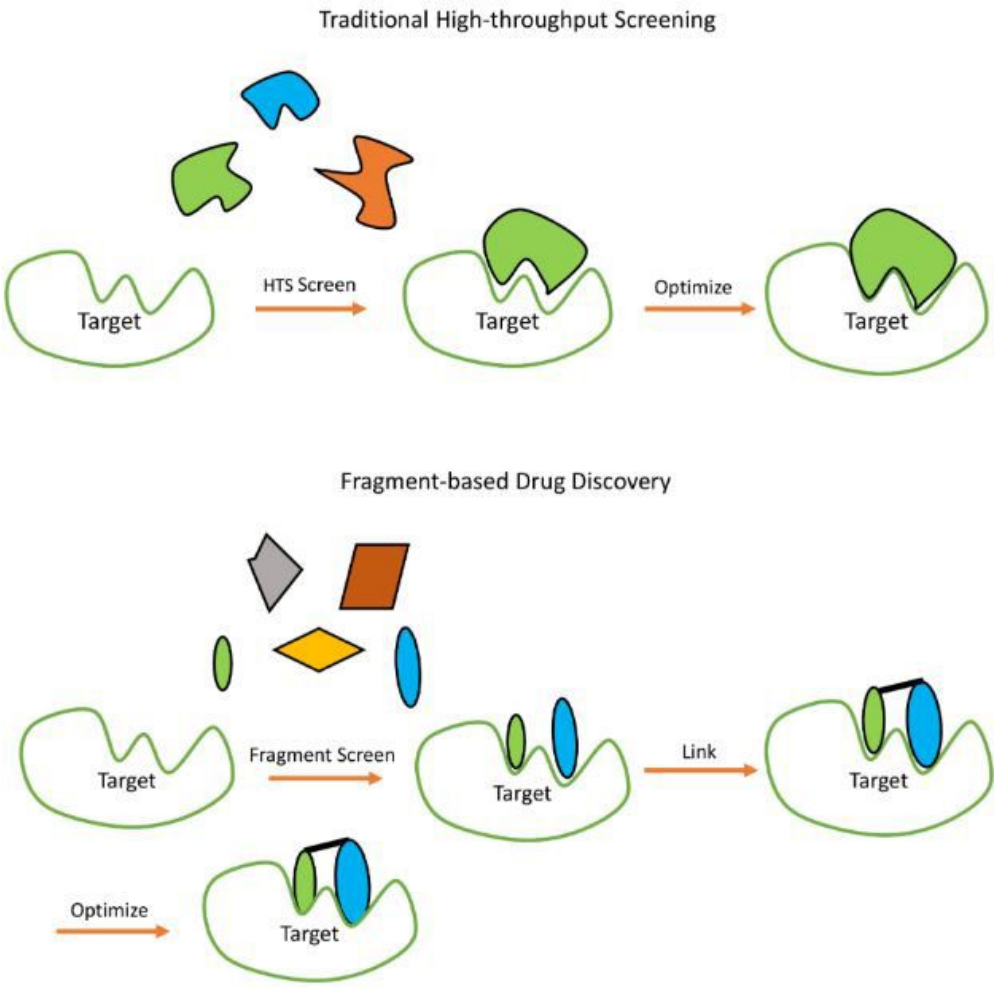
1.Reference: Neil Savage, "Tapping into the drug discovery potential of AI….",Biopharmadealmakers Feature, 2021,B37-B39, https://doi.org/10.1038/d43747-021-00045-7

Big Picture of Novel Chemical Entity Module



Fragment-based de novo design

Fragment-based de novo design (FBDD) is an attractive approach for early-stage drug discovery and development. The fragments are considered for developing compounds. We have an exclusive fragment library, which is diverse, and curated from the literature. The fragments in our library follow the rules of three in which fragments have a molecular weight of less than 300 Da, ClogP value of less than three, and less than three hydrogen donors and acceptors [2]. The individual fragments are docked to the binding pocket of the target and fragments showing good interactions are taken as seeds to build the rest of the compound.



2, Reference: Qingxin Li, "Application of Fragment-Based..", Front. Mol. Biosci.,2020,7:180 https://doi.org/10.3389/fmolb.2020.00180

Why Fragment-based drug design is advantageous to other techniques?

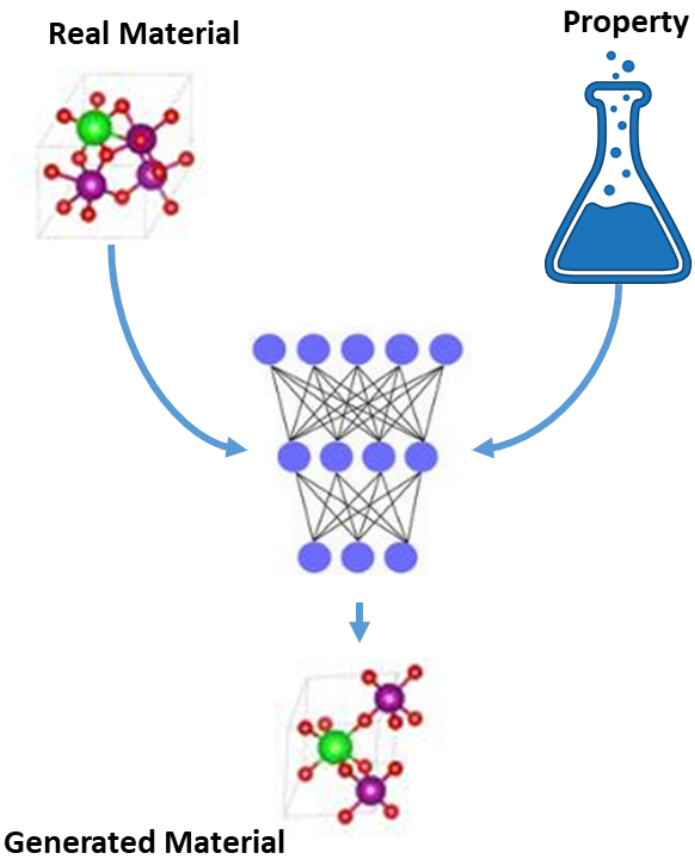
FBDD is mainly applicable to structure-based de novo design. In structure-based de novo design, the druggability of the target protein is analyzed. Target with defined pocket interacting with a drug is a druggable target. An undruggable target on the other hand does not have defined pockets or the pockets will be shallow. The undruggable targets include unstructured proteins which are crucial in many disease development and regulation. Traditional drug discovery projects do not consider these undruggable targets due to their high complexity nature. Studies show that FBDD can effectively be used to develop drugs for these undruggable targets [2].

This FBDD approach has striking advantages over conventional high-throughput screening campaigns. The high-throughput screening generally uses large libraries of complex compounds whereas, FBDD uses small libraries of low complexity fragments for screening to check for activity against a target. By using low complexity fragments, it is possible to locate more potential secondary binding sites of the target protein. Since only a small library of fragments is used, FBDD demands a lower investment in R&D than HTS. The molecular complexity is inversely proportional to binding probability, thus fragments with low complexity have good binding probability, which means that FBDD has high hits rate than HTS. Fragments being small entities can be iteratively optimized to generate compounds with better ligand efficiency and pharmacokinetic profile in the later development stages [3].

3. Reference: De Souza Neto LR, et al. “In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery”, 2020, Front. Chem. 8:93. doi: 10.3389/fchem.2020.00093

AI in De novo Design (AIDD)

In general practice, drug-like molecules that have activity against a particular target are screened from a huge library of compounds. It is still a challenge to pick synthesizable drug-like molecules, with desired molecular properties from the vast chemical libraries in a comparatively reasonable time. Instead of picking the drug candidate and then modifying its properties, it is wiser to produce drugs with the desired property. This is referred to as inverse design and it has been considered very difficult in the past. But, thanks to advances within the field of machine learning and artificial intelligence, the approach is becoming popular.



To design a molecule with the desired property, two neural networks work together. One is to generate the compound, generative model and the other is to predict and optimize the molecular properties, reinforcement learning-based multi-parameter optimization (RL-based MPO).

Generative Model

The generative model can create entirely new molecules, which is one of the most attractive benefits of AIDD. The generative model is composed of multiple layers of artificial neural networks. Based on the type of neural network used in the generative model, the input layers are fed with either SMILES string or a graphical representation of the molecule.

Initially, simple AI generative models using recurrent neural networks (RNN) are developed. The molecules are converted to SMILES string (sequence string representing format) and are treated as sequence strings. The RNN gets the sequence strings as inputs and generates special sequences as output. From the output sequence strings, the new molecules are developed. The method is not quite successful since the generated output sequence strings most of the time fail to develop a valid chemical structure. This is overcome by using DeepSMILES representation, which avoids problematic rings and brackets of regular SMILES to generate more meaningful sequences for developing new molecules character by character [4].

Further, the limitation of using sequence strings in the molecular generation is overcome by using advanced AI generative models such as variational autoencoders (VAE), generative adversarial networks (GAN), and reinforcement-based models. These models are considered promising for the computational creation of novel molecules due to their state-of-the-art results in the virtual synthesis of images, text, speech, and image captions.

RL-based MPO

In any drug discovery project, it is important to predict the properties of drugs such as absorption, distribution, metabolism, and excretion (ADME), toxicity, synthesizability, and many more. There are neural networks called multi-parameter optimizing (MPO) neural networks to accurately predict, filter, and improve the predicted properties of molecules before pre-clinical testing to reduce the cost of R&D.

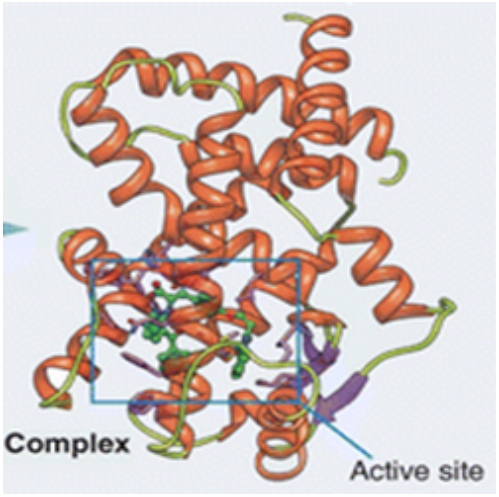
The multi-parameter optimizing neural network interacts with the molecules generated by the generative model, predicts its molecular properties, and modifies them to get desired properties. The generative model controls the RL-based MPO network. Thus by coupling the generative model and RL-based MPO we will be able to generate a new drug candidate with desired properties.

4. Reference: Niclas Ståhl, et al. "Deep Reinforcement Learning…", 2019, J. Chem. Inf. Model., 59:7, 3166–3176, https://doi.org/10.1021/acs.jcim.9b00325

Filtering & Hit Identification

Our module has molecular docking and ADMET property prediction tools to identify the best hit with good activity and drug-like properties, against a particular therapeutic target.

Molecular docking



Preparation of the input compound and the protein target is done before molecular docking. The input compound is docked to the binding site of the target. Docking is done to identify the best binding mode of compound showing good interactions with active site residues of the target, thereby having good binding energy and binding affinity values. The ranking is done based on the docking scores. The compounds which show favorable interactions and good docking scores are filtered out and carried forward for fur. Visual inspection of the docking results can further help the user to pick the best compound with favorable interaction for further development.

ADMET property prediction

Most drugs fail at clinical trials mainly due to lack of safety and because of their poor pharmacokinetic properties. Predicting ADMET properties covers these pharmacokinetic and toxicity issues. To avoid failures in clinical trials, nowadays the ADMET properties are predicted at an early stage of hit identification.

Our module uses multi-task graph neural networks to predict the ADMET properties. Most of the available free and commercial ADMET predicting tools are designed to predict either one or two properties simultaneously. Whereas our -

- multi-task graph neural networks can predict all the ADMET properties simultaneously and accurately. Apart from predicting the values of ADMET properties, our module also provides additional suggestions, which can be used to further optimize the molecular structure of the compound to convert them into a potent drug.

Hit Optimization

We use molecular dynamics (FEP) studies to optimize the hit.

FEP

We use the Free energy perturbation - thermodynamic integration method to accurately estimate delta G, the free energy of binding. We decouple the bound ligand in multiple intermediate lamda windows and estimate delta G change between ligand-bound and unbound to give the delta G of binding. Literature shows that this method has reported a root mean square error value of less than 1 kcal/mol. The absolute binding free energy (ΔG_{bind}) was calculated as $\Delta G_{bind} = \Delta G_{comp} - \Delta G_{lig}$.

FEP is successful in precise calculation of binding energy of protein-ligand complex guiding in the small molecule drug discovery.

Outcome

Our AI-driven de novo designing platform accelerates the drug discovery process. The use of a generative model coupled with RL-based MPO generates compounds with desired properties, thereby reducing the failures at a later stage of drug development. Further with other state of art techniques our module develops a potential, synthesizable, safe, novel chemical entity for a particular therapeutic target.

Why do you want to use our NCE module?

By choosing Medvolt's NCE module you take giant steps forward in the speed of finding novel chemical entities.

By developing compounds from scratch with desired pharmacokinetic properties, our module saves you a huge sum of money lost in failed trials.

Our user-friendly interface and domain experts will help you optimize the compound with ease, into a clinically safe drug.