AI toolkit for early-stage drug discovery

Learning by doing: Adopting AI one-step at a time

Nov 2020

Advisors | MEHTA PARTNERS

Table of contents

AI can catalyze the hit to lead optimization timelines

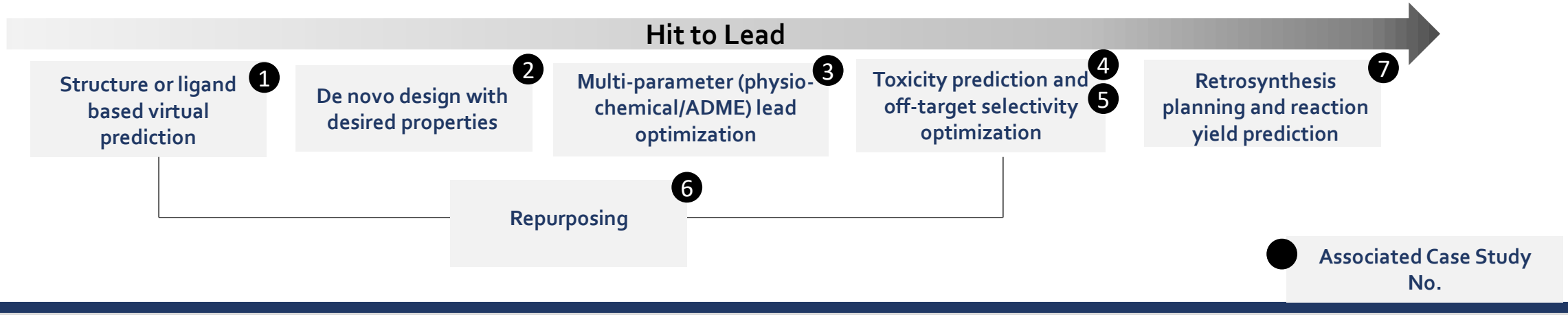
AI can 'learn' by analyzing big data and perform accurate predictions at an unmatched speed reducing costly manpower and experimental input

Sophisticated suite of computational drug design tools ranging from classical structure-based QSAR techniques to more recent advances in matched molecular pairs and free-energy perturbation methods have been used for about 2 decades in early drug discovery

Why the sudden hype about AI?

Recent widespread interest has been fueled by the breakthrough made in neural networks and generative adversarial networks (GANs) or often referred to as 'creative AI' in 2014. These algorithms can used to generate novel molecule entities with a desired set of pharmacological properties, efficiently predict toxicity, and many other applications in drug discovery

Why can AI do?



Why use AI?

AI offers rapid 'design-make-test' cycles at relatively low costs and fractional timelines to building a robust and diversified pipeline

Timelines: AI platforms can design potent lead-like molecules within 2-6 weeks and guide synthesis pathways for molecules, allowing rapid 'make-test-design' cycles as compared to 1-2 years timeline with traditional discovery methods

Cost: The platform predicted molecules often have micromolar/submicromolar range binding affinity during the first few cycles itself, therefore, the number of molecules (in 10-100s) required to be synthesized and tested are multi-fold lower as compared to traditional discovery (few thousands), leading to low cost

Robustness: Powerful ontology generation using data from systems and network biology to support evidence driven hypothesis generation enables the exploration of the hardest and the most difficult targets/programs

Accuracy: Better prediction power for features like solubility, ADME, synthesis and other features than traditional computational tools used, significantly reduces the chances of failure

Atomwise

1 Structure-based virtual screening

Uses AI to predict new active molecules for targets with no previously known modulators

AtomNet is deep convolution neural-network based algorithm for structure-based drug design and discovery. Its speed and accuracy makes it the most advanced technology for small molecule binding affinity. AtomNet has beaten autodock smina and DOCK on several benchmarking datasets

AI driven virtual screening and hit identification

Data pharma/biotech provides: Target information (Xray structure / homology model / protein-protein interaction information / cryo-EM structure / any previously known binders, if known)

What AI company provides: ~16 billion compounds for screening, platform for screening of molecule and 'hits' with optimal binding at the end of exercise

Process: Atomwise provides operations service and works closely to with partners to screen either Atomwise's library or client's proprietary library. Pharma will test the compounds for activity

When will it work: Reasonably accurate structure or binding information available.

When will it fail: Very low information available about the target or poor structural data

Benefit over traditional method: Faster and accurate Kd prediction. Average success rates of ~75% with turnaround time of hits around 1-2 months

Case study: Inhibitors against Ebola glycoprotein 2

AIM: In 2015, during Ebola outbreak, Atomwise partnered with University of Toronto and IBM to rapidly develop a treatment for Ebola virus infections.

SOLUTION: Atomwise defined a region to investigate for potential small molecule inhibitors by drawing a box around the bound outer helices in a closed x-ray crystal structure, and then deleted the helices, leaving the central core. Atomwise screened 7000 phase 2 or better compounds against Ebola virus glycoprotein 2 and identified a compound with no previously reported antiviral activities.

RESULT: Experimental validation evaluated a dose-response IC50 of approx. 12uM. This binding affinity is remarkably high given the molecules Atomwise screened were only those drugs that had been tested for safety and already pursued as treatments for other diseases. Molecules were identified within 2 months of effort.

Source: MP analysis, company website

Curr Pharm Des. 2018 Aug; 24(28): 3347–3358

Insilico Medicine

2 De novo design with desired properties

Uses AI to identify novel targets and design small molecules with desired properties

InSilico's platform (chemistry42) takes a chemistry-based (QSAR type or generative) approach for drug design against a single defined target. Their generative algorithm are driven by training on extensive chemical information

AI driven de novo design

Data pharma/biotech provides: Target information, known hits/ligands with associated activity, if available (increases Probability of success)

What AI company provides: AI platform trained on diverse data points and lead-like molecules optimized for activity and desired physiochemical and ADME properties

Process: Insilico either provides operations service or access to cloud-based platform

When will it work: The target should be well characterized with some information available about to ligands that can bind to the target

When will it fail: Very low information available about the target or no information on binding of ligands

Benefit over traditional method: Provides a molecule optimized for binding, activity, and desired ADMET parameters from the beginning of the process, leading to lower failure rates in subsequent steps. The total process of multiple cycles of AI generation to **novel lead-like molecule can be finished in weeks**

Source: MP analysis, company website

Case study: Inhibitors of DDR1, target for Fibrosis

AIM: Find patentable chemical ligands with new scaffolds for discoidin domain receptor 1 (DDR1), a tyrosine kinase target for which at least eight different chemotypes have been reported for it in the last few years

SOLUTION: Training the system with all the existing DDR1 literature, the larger set of kinase inhibitors in general, databases of medicinally active structures, and an even larger set (17,000) of compound structures that have been specifically claimed in all sorts of med-chem patents. The initial output was 30,000 structures. They cleared out structures based on molecular weight, number of polar groups, etc., and unstable or reactive molecules. Clustering and chemical diversity sorting reduced it to 4600

Applying kinase-evaluating filters and fitting to pharmacophore models derived from known DDR1 ligand-bound X-ray structures reduced it to 848 candidates, and the group then picked 40 structures that scattered across the chemical space.

RESULT: Molecules were entirely outside of currently patented chemical matter. Team chose six of them for experimental validation, two of them hit DDR1 at around 10 and 20 nM, two others were up in the hundreds of nanomolar range

The result were validated with in-vivo testing within 46 days

Source: Nature Biotechnology volume 37, pages1038–1040(2019)

IKTOS

3 Multi-parameter lead optimization

Uses AI to optimize and design small molecules with desired properties

Iktos platform can create a customized multi-parameter QSAR filter to screen and optimize the leads. They need data from the partner to optimize the outcome of the algorithm

AI driven lead optimization

Data pharma/biotech provides: Known hits/ligands with associated activity and other assay data, if available (increases Probability of success). Otherwise, data can be extracted from publicly available sources

What AI company provides: Customized multi-parameter QSAR model and optimized lead-like molecules

Process: Iktos provides operations service and pharma tests the molecule

When will it work: The platform is trained on very diverse data and has data for molecules somewhat similar to being evaluated

When will it fail: if training data is very different from the molecule under consideration

Benefit over Traditional method: Provides simultaneous optimization of lead for desired ADME and known off target parameters leading to reduction in number of compounds required to be synthesized and tested. If relevant data (see case study) is available, the whole process can be finished in 2-3 months

Case study: optimized lead

AIM: : Identify molecules meeting 11 objectives of the project simultaneously. The project data set comprised 880 molecules tested on 11 biological assays: 1 activity criteria (phenotypic assay), 6 off-target activity (selectivity criteria), 4 DMPK criteria (microsomal stability and permeability assays). The data set was sparse with 10-70% missing data rates. 6 active molecules were meeting a maximum of 9 objectives

SOLUTION: Iktos DL- based de novo design algorithm was used to optimize leads and design virtual molecules fulfilling all 11 objectives with a proprietary multi-objective fitness function built from the individual predictive QSAR models

RESULT: 150 virtual compounds meeting the criteria were identified. 20 compounds were selected based on synthetic accessibility, structural diversity, and score confidence. The algorithm was able to suggest functional groups that were rare or absent in the initial dataset. For 9 molecules the synthesis failed, so 11 compounds were finally tested. The 11 AI designed molecules averaged 9.5/11 objectives.

One molecule met all 11 objectives

Source: MP analysis, company website

CYCLICA

4 Off-target toxicity prediction

Uses AI to consider a polypharmacology approach to drug design and optimize for high selectivity

Cyclica, unlike most AI companies, takes a protein structure centric approach for drug design against multiple targets at the same time. Their drug design is powered by a rapid deep learning based whole proteome screening approach followed by chemistry driven optimization

AI driven toxicity prediction-structural approach

Data pharma/biotech provides: Known hits/ligands for which targets are unknown or desired target

What AI company provides: Targets identification and optimized leads

Process: Cyclica provides operations service or/and access to subscription-based platform and pharma tests the molecule

When will it work: The platform can run on full generative mode and will likely create leads without any additional target specific training

When will it fail: Chances of failure if the structural model of protein is very bad or homology info is poor

Benefit over Traditional method: Provides a simultaneous multi-target (desired on and off-targets) approach for high selectivity pressure, leading to optimized binding and reduction in any off-target toxicity, thus higher probability of success in clinic

Case study: Uncovering potentially fatal drug-protein interactions

PROBLEM: Fatty acid amine hydrolase (FAAH) inhibitor BIA 10-2474 demonstrated severe neurological trauma in clinical investigations, leading to death of one patient and hospitalizations of five others

SOLUTION: Cyclica's AI platform Ligand Express screened the molecule against the whole proteome and generated a complete list of proteins that are likely to interact with BIA 10-2474 (the desired target, FAAH, was identified as the 104th hit, top 99.9 percentile)

The list was restricted to proteins that were related to the blood and brain using Network Analysis and systems biology

RESULT: Ligand Express predicted binding of BIA 10-2474 to coagulation factor VII (FA7) in the same region as a known inhibitor. Inhibition of coagulation factor VII by BIA 10-2474 may cause deadly neurotoxicity (bleeding within the brain) as observed in the French clinical trial

Source: MP analysis, company website

CYCLICA

5 Off-target selectivity optimization

Uses AI-based polypharmacology approach to drug design and optimize for high selectivity

Cyclica, unlike most AI companies, takes a protein structure centric approach for drug design against a multiple targets at the same time. Their drug design is powered by a rapid deep learning based whole proteome screening approach followed by chemistry driven optimization

AI driven multi-target drug design

Data pharma/biotech provides: Known hits/ligands for which targets are unknown or desired target

What AI company provides: Targets identification and optimized leads

Process: Cyclica provides operations service or/and access to subscription-based platform and pharma tests the molecule

When will it work: The platform can run on full generative mode and will likely create leads without any additional target specific training

When will it fail: Chances of failure if the structural model of protein is very bad or homology info is poor

Benefit over Traditional method: Provides a simultaneous multi-target (desired on and off-targets) approach for high selectivity pressure, leading to optimized binding and reduction in any off-target toxicity, thus higher probability of success in clinic

A Multi-target design approach

Cyclica used its platform to identify binders to the panel of 5 proteins the protein deacetylases sirtuin 1, 2, 3, the apoptosis regulator MCL1, and the non-receptor protein tyrosine phosphatase (SIRT1, SIRT2, SIRT3, PTP1B, MCL1) to design in Sirtuin 2 inhibitor. Only about 100,000 compounds were screened of which 22 were selected and Eight of the 22 molecules showed measurable inhibition for at least one target.

Molecules were screened, purchased, and tested in vitro within 20 business days.

CYC-1858 demonstrates activity at low micromolar concentrations and represents a novel scaffold for the selective inhibition of sirtuin 2. The determined IC50 of 2.0 ± 0.1 µM is comparable to current best-in-class selective sirtuin 2 inhibitors

Source: MP analysis, company website

PHARNEXT

Uses AI to screen and reposition known drugs in unrelated indications at new, lower doses

Pharnext's platform processes multi-omics data and drug-disease mapping data to identify a suitable combination of targets for the indication and identify drug synergies/combinations that can act on the multitude of targets identified

AI driven repurposing

Data pharma/biotech provides: Molecules that need to be repositioned, any associated phenotypic data, or an indication for repurposing

What AI company provides: Operations service or platform subscription to generate insights

Process: Both parties work together

When will it work: The platform is trained on a diverse amount of phenotypic/OMICs data

When will it fail: Very little information about the indication is known

Benefit over Traditional method: Provides quick, robust, evidence-based insights (with references) to guide the repositioning and repurposing

PLEOTHERAPY™-repositioning drugs for a rare disease

Pharnext reported positive results for a Phase III trial of one of its drug combinations, PXT3003 for a neurodegenerative condition called Charcot-Marie-Tooth disease (CMT), a rare disorder.

PXT3003 is a low-dose fixed combination of baclofen, naltrexone and sorbitol, given twice a day as an oral solution. It has multiple mechanisms of action - inhibition of PMP22 gene associated with an improvement in myelination, preservation of the axon of the peripheral nerves and additional targets as identified by systems biology approaches

PTX3003 was granted orphan drug status by US FDA and EMA

Source: MP analysis, company website

Molecule.one

AI in Chemical synthesis reaction pathway, yield and condition prediction

Uses AI heavily trained on curated synthesis reaction data to identify stereochemical orientations during reaction and predict the feasibility of synthesis for a given chemical reaction/compound

AI driven chemistry optimization

Data pharma/biotech provides: Structure of molecules that need to be synthesized

What AI company provides: Synthesis pathway prediction, or if pathway is known optimal reaction conditions for high yield

Process: AI company provides operations service and sometimes the synthesized molecule

When will it work: The platform is trained on a diverse amount of chemical reactions

When will it fail: Having metals or very heavy atoms in the structure

Benefit over Traditional method: Provides quick insights into chemical synthesis pathway and optimize the yield, hence improving the timelines and decreasing cost to the end product

Merck KGaA compound synthesizability challenge

Molecule one participated in in Merck's challenge for predicting feasibility of a particular reaction.

They trained their proprietary algorithm with data that consists of 400 000 known reactions scraped from publicly available patents and grouped the reactions into 1025 templates (reaction types) based on the characteristics of the neighborhood of the reaction center. They also included a total of 40 000 negative samples - reactions that we treat as infeasible.

The algorithm can predict reaction feasibility with an accuracy of above 90%

Source: MP analysis, company website

Key considerations for evaluating AI algorithms

Data

Training data: Understanding the number of training points and diversity of chemical structures and associated information (LogP, LogS, number of ADMET parameters) used is important to check for relevance of platform to the specific project requirements.

Data needed: Some applications (like activity prediction) may need specific data from the pharma/biotech for the platform to efficiently predict the values

Models/Algorithms and Performance Characteristics

Type of ML/DL algorithm: Different kinds of algorithms (RF, SVM, Bayesian, CNN, RNN etc) perform differently across the diverse steps in drug discovery. For e.g. CNN is great for structure-based approach but poor in predicting ADME, whereas SVM or Bayesian models perform the best for the latter. Featurization approach also affects the performance between two models and should be taken into consideration during the selection.

It is critical to assess the performance characteristics/values of the platform along with the depth of testing data while selecting the platform

Approach

Ligand centric vs. Structure based: AI platforms can approach drug design through ligand and Structure centric approaches. Each approach has different prerequisites for success and should be a deciding factor in selecting the platform's applicability

Transfer learning: Adding insights extracted from text search from patents, publications, etc. (using Natural Language Processing or NLP) and systems/network biology approaches can improve the accuracy and efficiency of the drug discovery applications. Only a few platforms have this capability

Benchmarking

While an algorithm may have high performance values on its Training/Test dataset, its performance may differ significantly on other datasets. Establishing an internal benchmark to compare different platforms for a variety of applications can aid in the right choice for adoption of AI

Source: MP Analysis

A framework to build a robust AI strategy

Step 1

Internal Foundation

Identify the area of focus

- Understanding deeply the chemistry segments that can benefit from AI
- Identifying internal segments for AI adoption with minimal investment
- Exploring potential opportunities through partnerships, investments or licensing platforms

Informed Team

- Assess comfort level of scientists about preference of platforms and adoption of different platform vs. integrating into current platforms
- Identify cross-functional teams and ensure continuous evolution of their understanding

Monitoring and Security

- Establishing key parameters and performance indicators for monitoring growth
- Data sharing in Biopharma is often a concern. Understanding data security and partners data is important

Step 2

Finding the Right Partner/Platform – Key Parameters to Consider

Therapeutic area knowledge (Domain specific or service specific)	Platform capabilities (Stringent or Adaptable to diff. data)	Management and team (AI experts vs.chemistry experts)
Proof of concept/TRL level (publications, patents)	Data used for training algorithm (depth of data, proprietary vs. public)	Business models (Platform for use or partnerships)
Other Partnerships (other big pharma, consortiums)	Scalability and evolution (Future generations/improvement possible?)	Cost Benefit (Tangible and intangible returns)

Source: MP Analysis

MP Group can catalyze your AI initiative

With over 3 decades of diverse experience and integrated perspective in domestic and global BioPharma, and deep understanding of AI space, MP Group has the capabilities to help establish your AI initiative

MP Team will be happy to be an extension of the management team and help with one or more of the below initiatives:

- Asses the internal capabilities and identify the key business segments for potential disruption/augmentation by AI platforms
- Identify business segments for short-term and long-term benefit from AI interventions
- Identify partnering or investment opportunities unique to your vision
- Assist with designing the initial feasibility studies
- Technical due diligence to investigate the AI platforms best suited for the need

THANK YOU.

We invite you to write to us -

Viren Mehta
mehta@mpglobal.com

Neel Fofaria
neel@mpadvisor.com

Amandeep Singh
amandeep@mpadvisor.com