Computer-Aided Drug Discovery

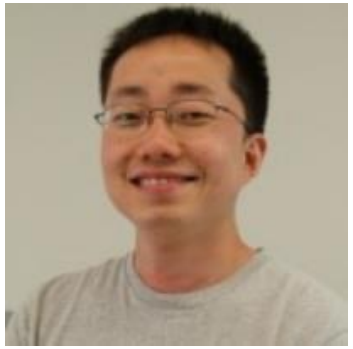
Seminar #2 - Butler Summer Trainee 2022

Pinyi Lu

June 30, 2022

Pinyi Lu

Senior Data Scientist, Cancer Data Science Initiatives, FNLCR



pinyi.lu@nih.gov

My Experience

- 16+ years research experience on translational informatics
- Collaboration with pharmaceutical companies and healthcare systems
- Recent works include:
  - JDACS4C Pilot 1 – Predictive Modeling for Pre-Clinical Screening
  - AI-based drug discovery and design of anticancer drugs that target CENP-E

Education

- Post-doctoral Appointment, University of Texas Health Science Center
- Ph.D., Computational Biology and Chemoinformatics, Virginia Tech
- M.S., Computer Science, Virginia Tech

Keywords

- CADD and Machine Learning
- Precision Medicine and Translational Research
- Badminton and Pickleball

JDACS4C: Joint Design of Advanced Computing Solutions for Cancer;
CADD: Computer-Aided Drug Discovery

Seminar #2 by Dr. Pinyi Lu

Date - Thu June 30, 2022, 11:00 am – 12:00 pm ET

Meeting Link - https://teams.microsoft.com/l/meetup-join/19%3ameeting_YjRiZGU5NzMtOTdlZC00OTQ5LTk2YWEtNzAzNmIwMmUwZWI5%40thread.v2/0?context=%7b%22Tid%22%3a%2214b77578-9773-42d5-8507-251ca2dc2b06%22%2c%22Oid%22%3a%222e31b4fb-2c54-4995-be48-173a24bc5b84%22%7d

TITLE - Computer-Aided Drug Discovery

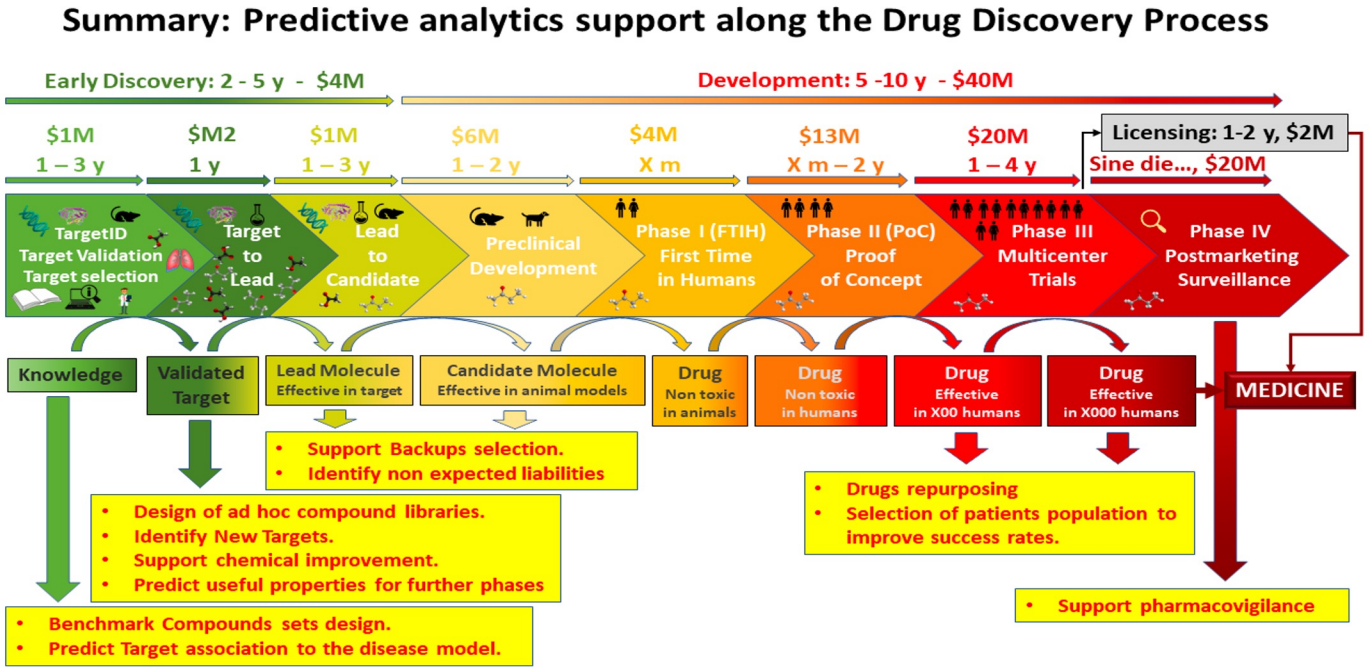
PRESENTER – Pinyi Lu, Ph.D., Senior Data Scientist, FNLCR

ABSTRACT - Computer-aided drug discovery (CADD) has become an important part of the drug discovery process due to the reduced cost of computational methods and the increased availability of three-dimensional structural information. In this seminar, an introduction to CADD will be addressed. I will compare structure-based and ligand-based modeling, focusing primarily on molecular docking and quantitative structure-activity relationship modeling. In addition, I will present an example of application of computational methods in drug discovery and highlight some considerations in the application of CADD.

Part I
Introduction to Computer-Aided Drug Discovery

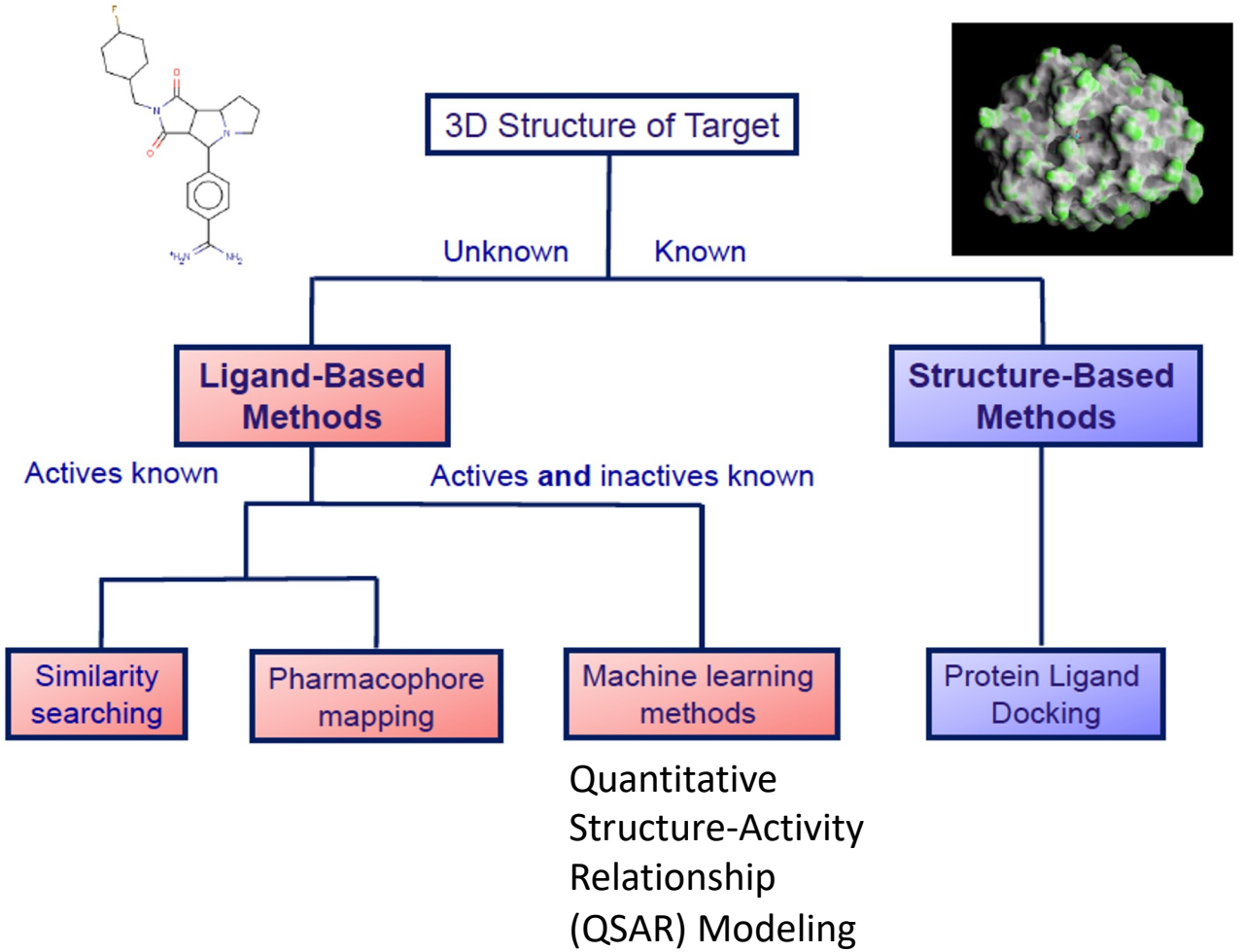
The Drug Discovery Challenge

- “Drug like” chemical matters, estimated between 10^22 and 10^60 unique molecules, evaluating them is a time-consuming and expensive process
- Computer-aided drug discovery, able to save cost, as well as generate a complete traceable and reproducible process



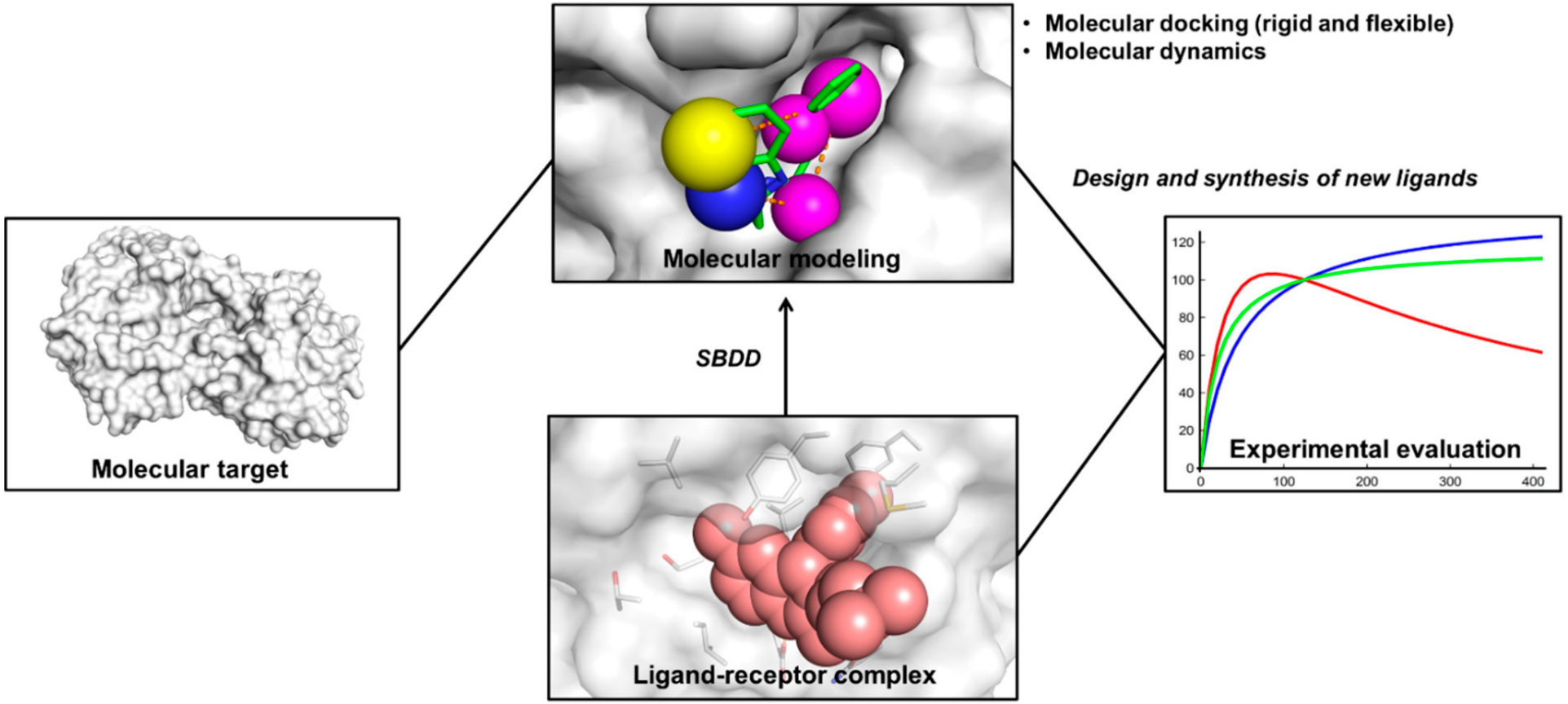
Source: Zhao L. et al. Drug Discov Today (2020); https://doctortarget.com/machine-learning-applied-drug-discovery/

Computer-Aided Drug Discovery



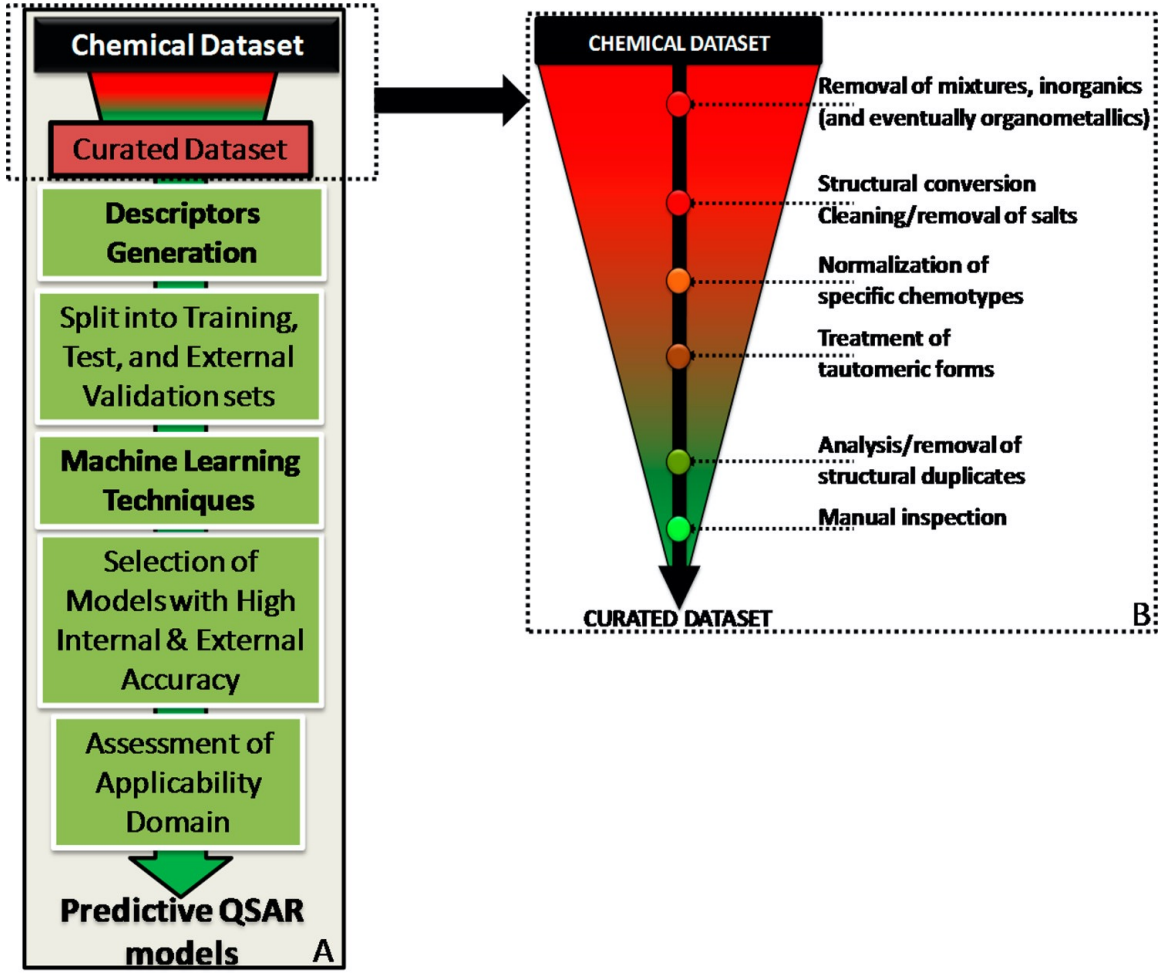
Source: Gillet V. Ligand-Based and Structure-Based Virtual Screening (2013)

Structure-based Modeling – Molecular Docking



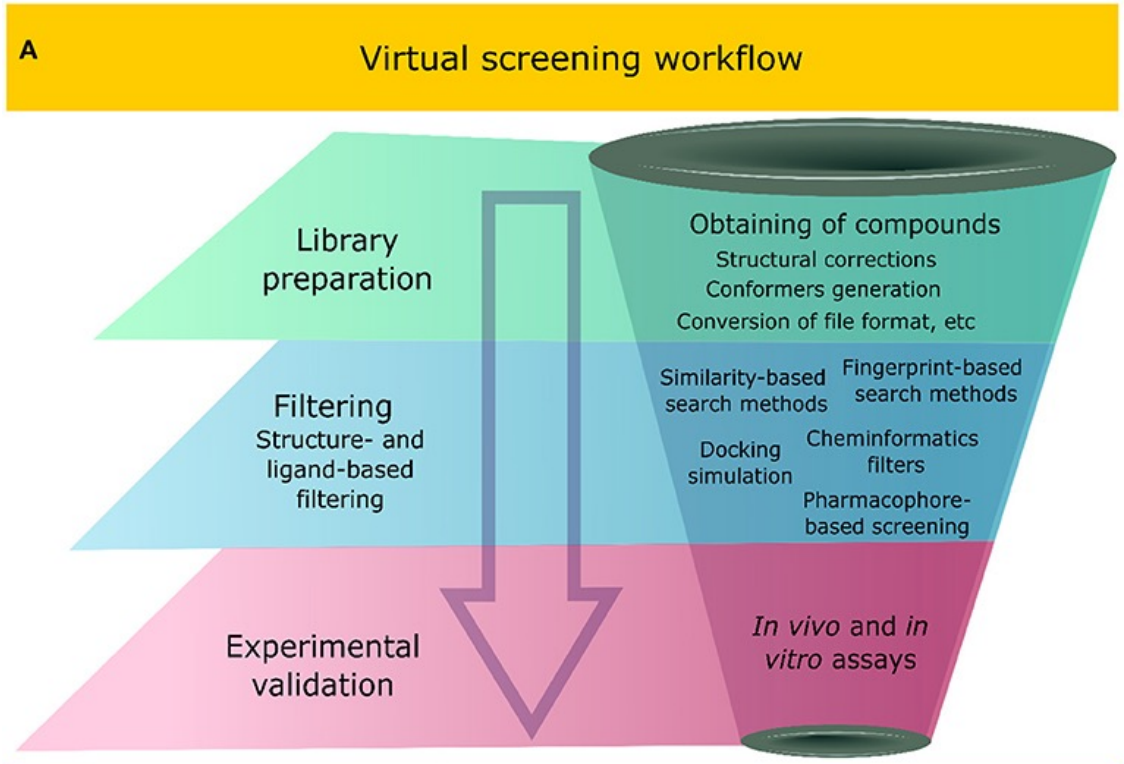
Source: Ferreira L. et al. Molecules (2015)

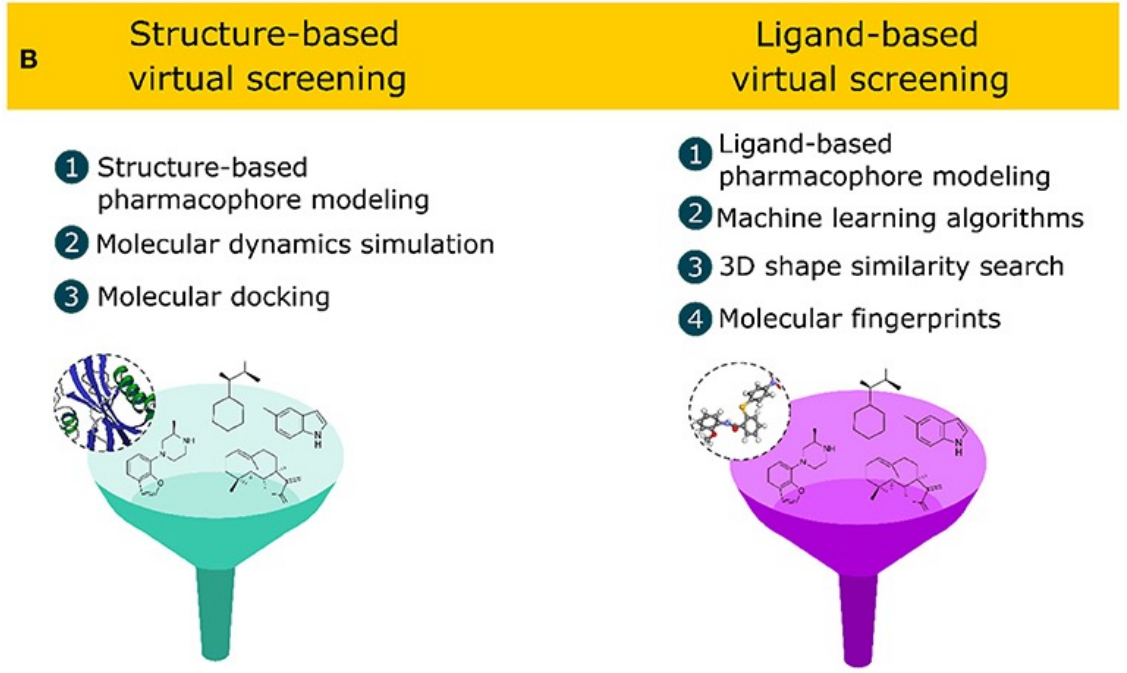
Ligand-based Modeling – QSAR



Source: Cherkasov A. et al. J Med Chem (2014)

Virtual Screening

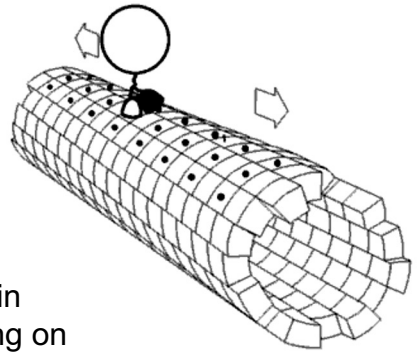




Source: Santana K. et al. Front Chem (2021)

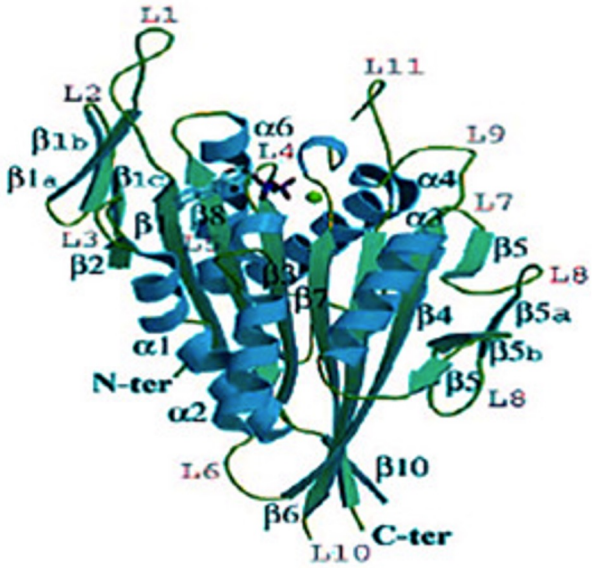
Part II
Example of Application of CADD

Centromere-associated protein-E



Kinesin
walking on
a microtubule

- Centromere-associated protein-E (CENP-E) is a mitotic spindle motor protein and inhibition of CENP-E is promising for cancer therapies
- CENP-E inhibitors could activate innate immune pathway and has potential to induce the immunological conversion from cold to hot in cancer cells
- None went beyond phase I clinical trials
- Not enough bioactivity data



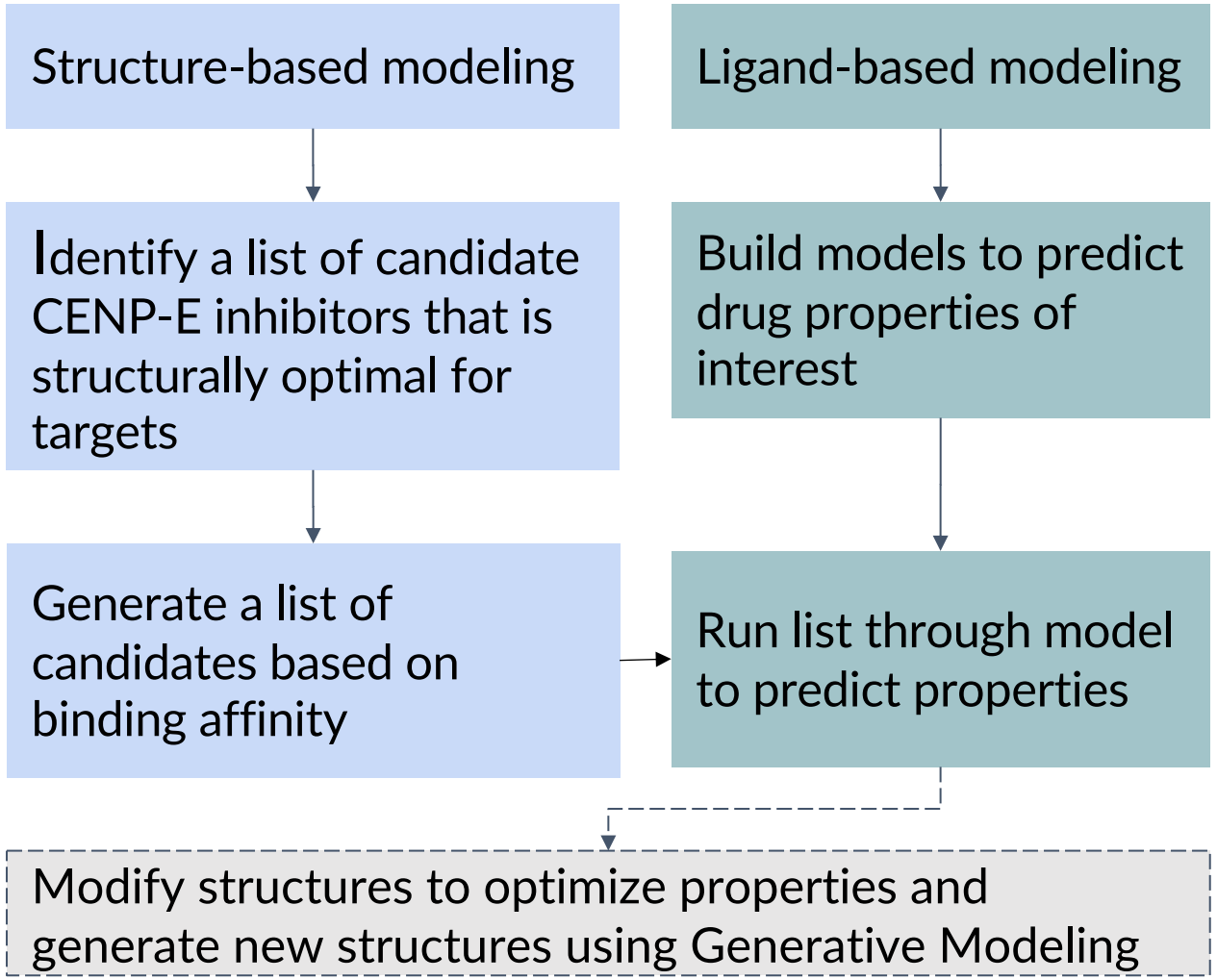
CENP-E structure

Source: Ohashi A et al. Nat Commun (2015); AhmedEl-Arabey A et al. Life Sci (2018); https://en.wikipedia.org/wiki/Motor_protein

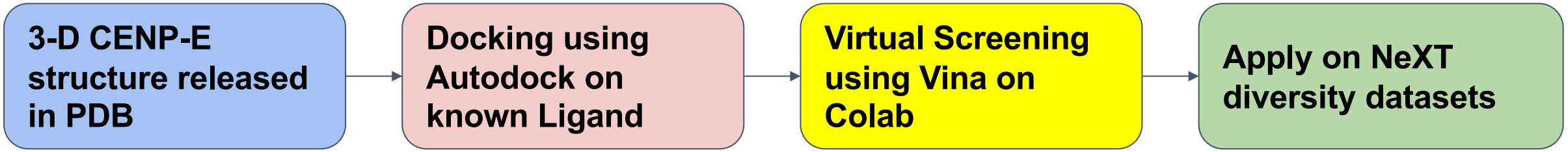
Methodology Overview

Objective

Develop a novel workflow integrating structure-based and ligand-based modeling approaches and apply this workflow to aid the discovery and design of novel CENP-E inhibitors



Methodology – Structure-based modeling on CENP-E



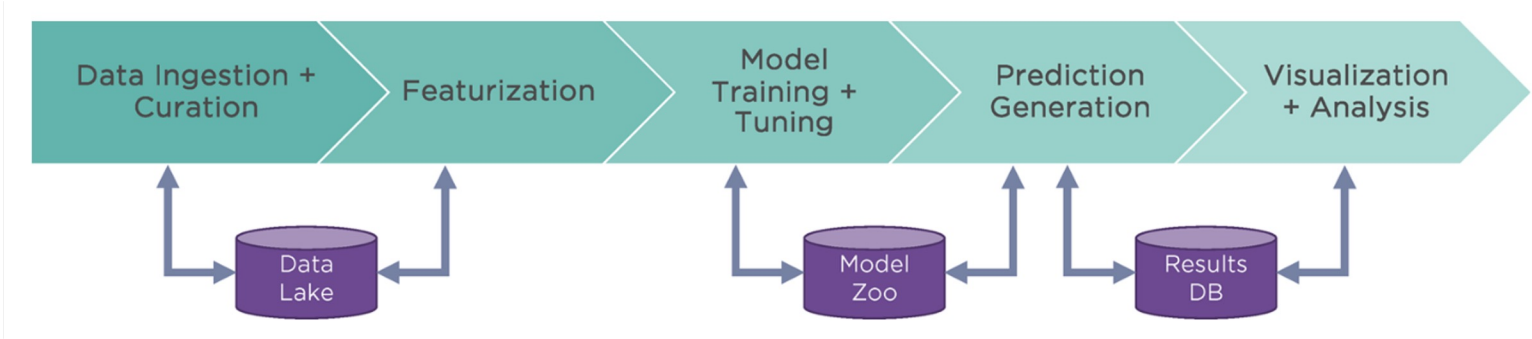
- Download 6m4i on Protein Data Bank (PDB)
- Further refined the protein structure using Chimera

- Perform Docking using Autodock 4.2.6
- Used known inhibitors, Compound-A and GSK923295 as references

- Used Package: Autodock Vina
- Auto-Docking on Google Colab

- Result extracted based on free energy of binding (binding affinity)

Methodology – Ligand-based modeling: AMPL

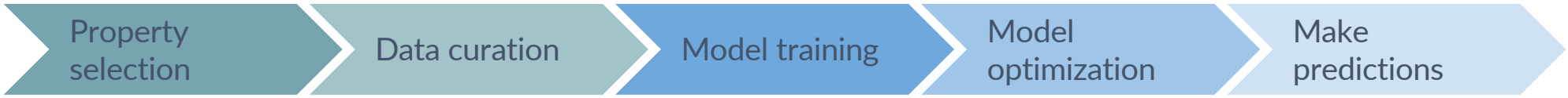


ATOM Modeling Pipeline (AMPL), extends the functionality of the open source library DeepChem and supports an array of ML and molecular featurization

- AMPL supports: Random Forest, XGBoost, Fully Connected Neural Network, Graph Convolutional Neural Network
- Featurizers: Extended connectivity fingerprints (ECFP); graph convolution latent vectors; Mordred open source package; Commercial software package Molecular Operating Environment (MOE)

Source: Minnich AJ et al. J Chem Inf Model (2020)

Methodology – Prediction on side effects



- **Chose side effects** "neuropathy" to predict using data from SIDER
- **Modify** the process on how to classify into neuropathy category

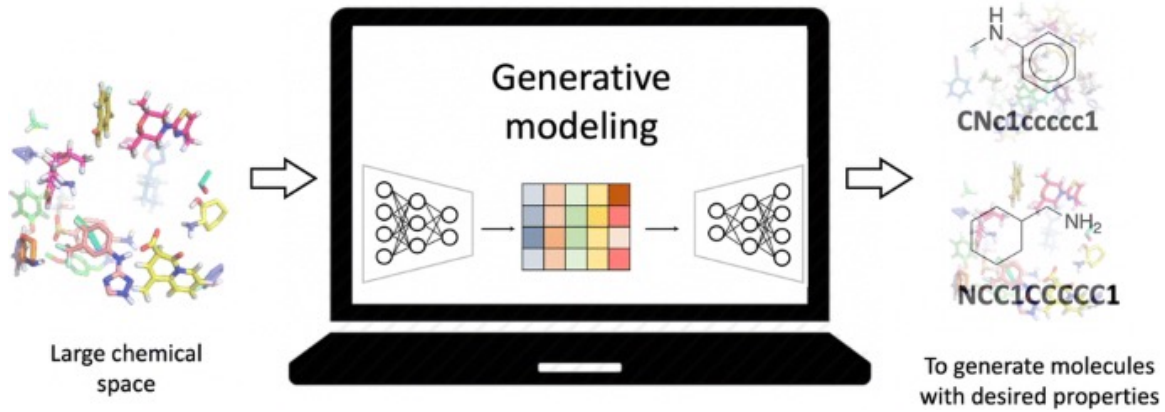
- **Obtained SMILES structure** by mapping SIDER data to PubChem
- **Classified data** into binary classification in respect to with/without side effect "neuropathy"

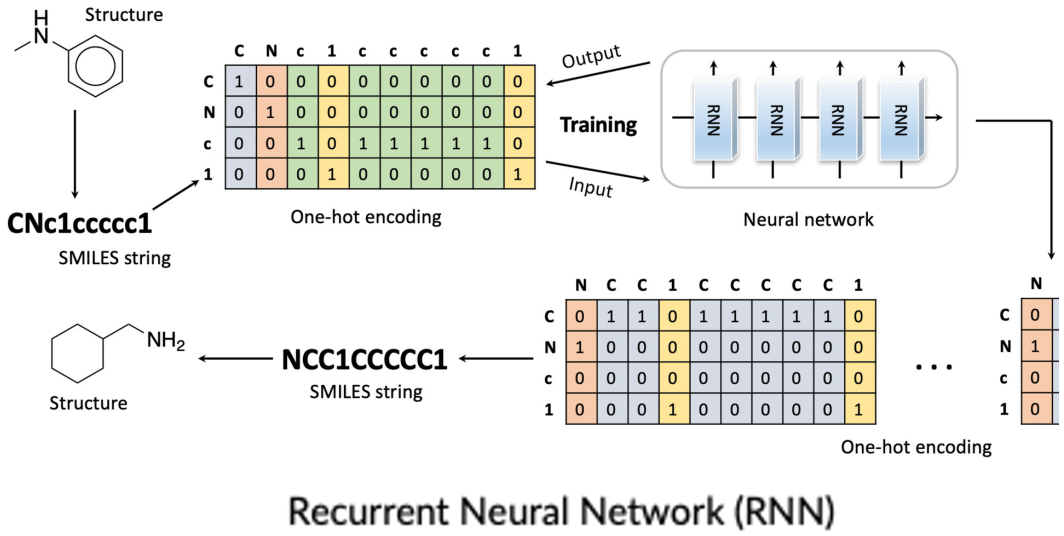
- **Leveraged Random Forest (RF) and Graph Convolutional Network (GCN) packaged in AMPL** to train model to make prediction

- **Performed balanced sampling** and repeated cross validation to improve prediction
- **Performed hyperparameter tuning** on both RF and GCN to optimize test result of ROC score

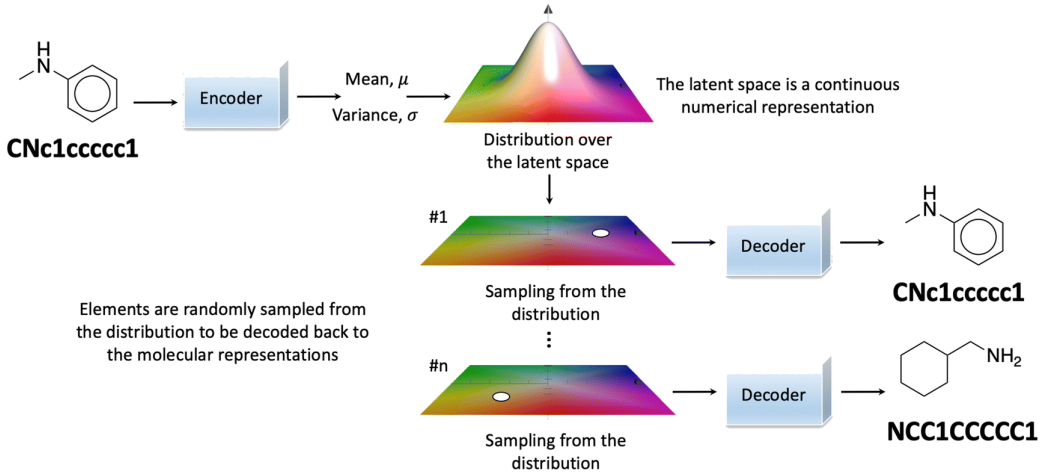
- **Make prediction on top-ranked compounds generated from structure-based modeling** which is the NExT dataset with binding affinity result

Methodology – Generative Modeling





Recurrent Neural Network (RNN)



Variational Autoencoders (VAE)

Source: Bian Y et al. J Mol Model (2021)

Conclusion

- Identified ligands which have high binding affinity with CENP-E and are favorable in terms of “Neuropathy” side effect
- Performed structure-based and ligand-based modeling to predict free energy of binding and side effect of potential CENP-E ligands
- Performed hyperparameter optimization to improve AMPL model performance
- Applied generative modeling to generate new structures of ligands

Acknowledgements

- Columbia University
- Frederick National Laboratory for Cancer Research
- Accelerating Therapeutics for Opportunities in Medicine (ATOM) Consortium

Thank You!