Blood Brain Barrier Permeability

Bioinformatics Society and TJML

Drug Discovery

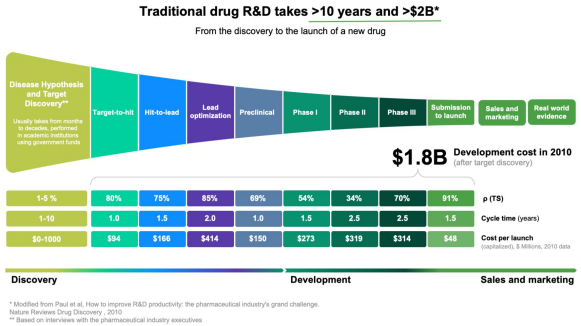
- The process of discovering new medicine to treat disorders
- Predicting properties of molecules
  - Toxicity, blood brain barrier permeability, determining IC50 concentrations, etc.

Current Approaches

1. Identify a critical target protein that modulates a disease
    a. Requires bioinformatics and expression analysis
2. Screen for molecules that can inhibit the target protein
    a. Requires combinatorial chemistry, structure-based drug design, and in vitro screening
3. Optimization
    a. Requires traditional medicinal chemistry & rational drug design
4. Evaluate pharmacokinetic properties
    a. effects on absorption, distribution, metabolism, and excretion
5. Clinical trials
6. Government registration
    a. FDA approval

Problems

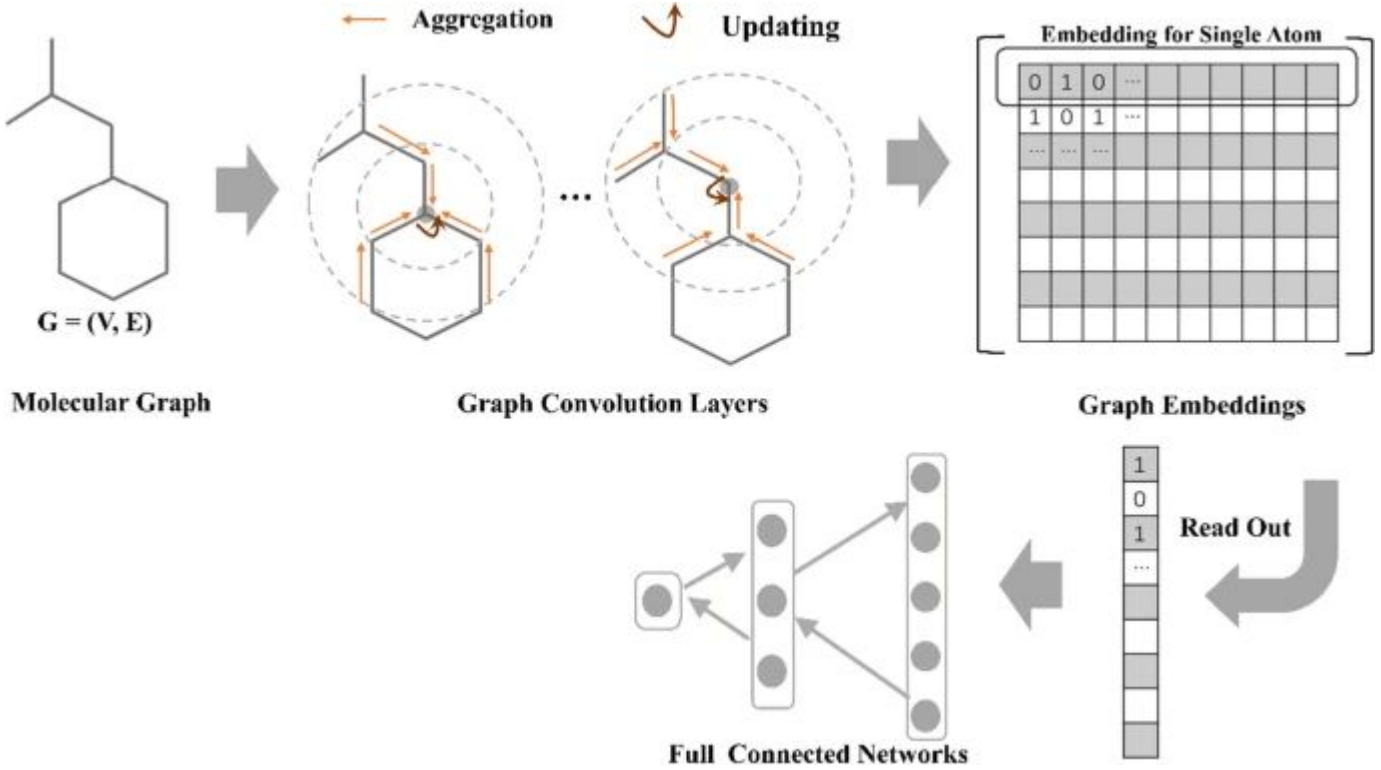
- Current drug discovery approaches very expensive and time-consuming
  - Costs upwards of 2 billion dollars and can take a decade of research and development
  - Trial and error
  - 15% of drugs make it through clinical trials



ML-Based Approaches

- Extract features from molecules to better predict how they interact with the body (toxic, permeable, can bind to specific proteins)
- Predict properties of discovered molecules
- Although prior research extracts features from molecules, current research focuses on learning molecule representations using neural networks
    - Graph neural networks, variational autoencoders

Graph Neural Networks

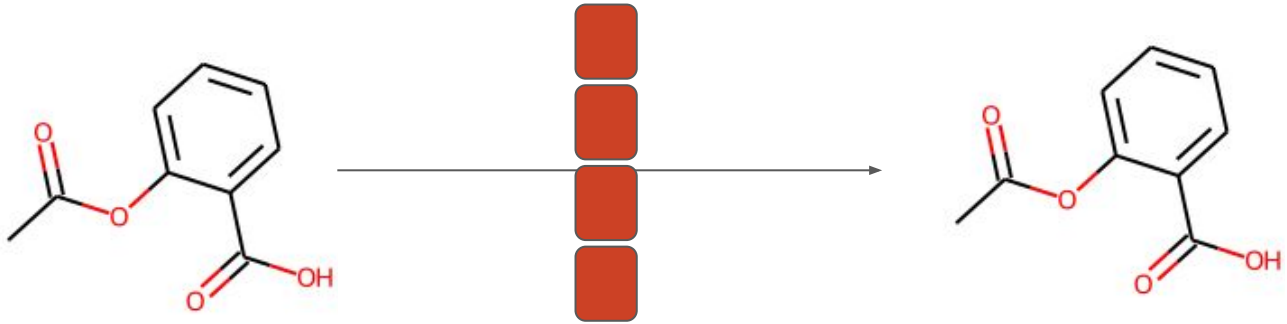


Implementation Details

- DeepChem library has open source implementations for many neural network architectures tailored for learning representation of molecule datasets
- Abstracts complex implementation details
  - Molecules have more constraints than typical nodes in graphs

Lab Objective

Determine whether a certain molecule/drug can permeate the blood brain barrier



Data Format

- CSV File → 3 Columns
  - Drug Name
  - SMILES: Simplified molecular-input line-entry system
  - OUTPUT: Permeable or Non-Permeable

name	p_np	smiles
Propanolol	1	[Cl].CC(C)NCC(O)COc1cccc2ccccc12
Terbutylchlorambucil	1	C(=O)(OC(C)(C)C)CCCc1ccc(cc1)N(CCCl)CCCl
40730	1	c12c3c(N4CCN(C)CC4)c(F)cc1c(c(C(O)=O)cn2C(C)CO...
24	1	C1CCN(CC1)Cc1cccc(c1)OCCCNC(=O)C
cloxacillin	1	Cc1onc(c2ccccc2Cl)c1C(=O)N[C@H]3[C@H]4SC(C)(C)...

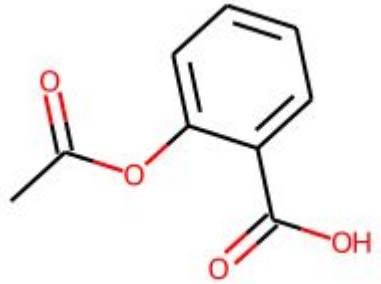
Drug Representation - SMILES

- Single line representation of molecular structure
- Commonly used in bioinformatics
- ASCII strings with characters representing bonds
  - Example: “=” represents double bond

SMILES Representation

`CC(=O)Oc1ccccc1C(O)=O`

Structural Representation



Step-by-Step Process

1. Load in the data
2. Use DeepChem to represent SMILE molecule as adjacency list
3. Initialize GNN model
4. Train Model
5. Evaluate

ACCESS LAB: tjmachinelearning.com → Lectures → Advanced