FL94 Capstone:

Super Glue – Interactive Web Tool to Facilitate Chemistry <> ML Collaboration

Team Members: Adrian He, Crystal Liu, Zoe Williams, Eric Ying
Project Lead: Dr. Orion Dollar (Flagship Labs 94)
March 18th, 2025

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

CHEMICAL ENGINEERING
UNIVERSITY of WASHINGTON

Flagship Pioneering

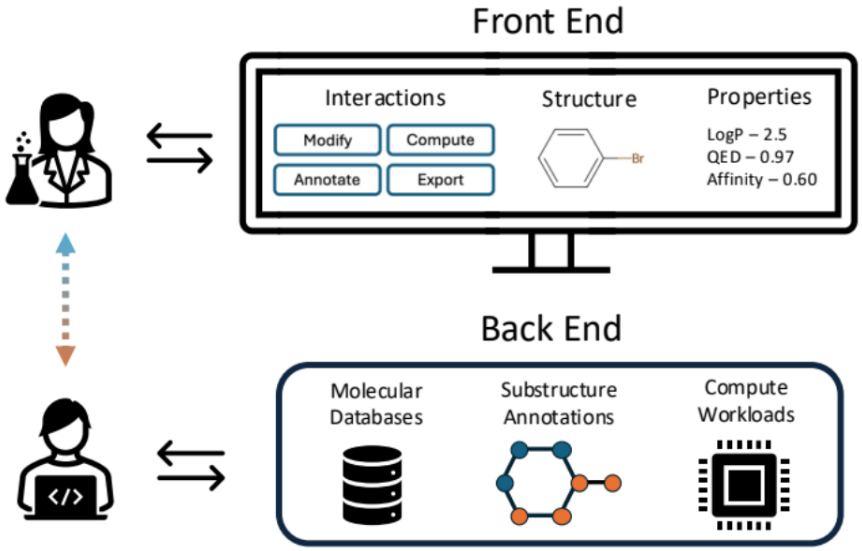
Agenda

- Introduction
- Design
- Technologies
- What we wished we knew earlier
- Moving forward

Introduction

Motivations

- Implement machine learning to pharmaceutical chemistry field
- “Glue” together expert experience and ML algorithms
  - Improve performance of models
  - Build trust along domains
  - Integrate ML into drug discovery pipelines
- Goal: Create an interactive tool to facilitate the collaboration of expert knowledge with ML in a scalable and data-driven manner



Introduction

Minimum Viable Product

- Upload .csv file containing molecule data and read into backend
- Display chemical structure through open-source chemical editing tool – Ketcher
- Display table to corresponding properties of molecular structures
- Allow users to annotate structures
  - Highlight substructures
  - Add text-based annotations
- Submit annotations as structured graph data to backend
  - Automatically written to disk
- Allow users to page through all molecules

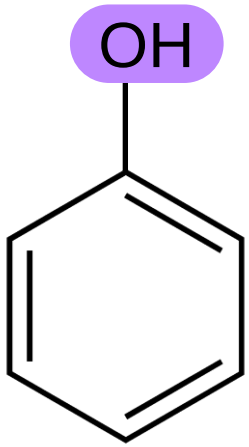
Design

Design

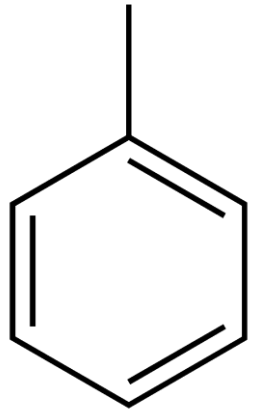
User Stories

csv dataset

highlight key
functional groups



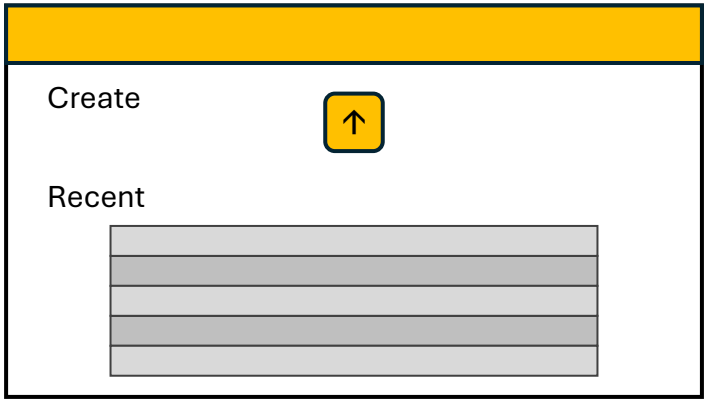
visualize molecule



find similar compounds

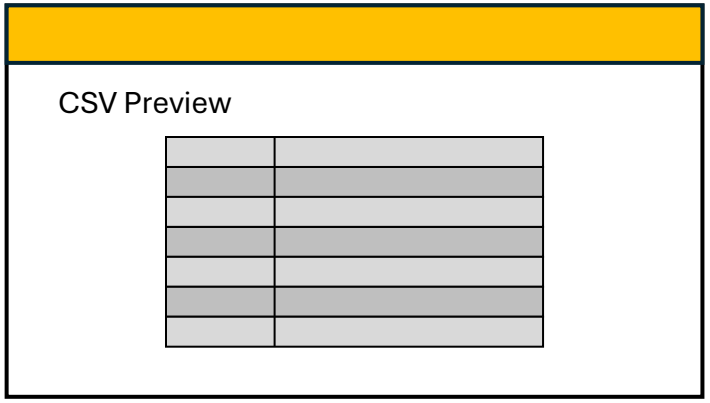
Design

Home Page



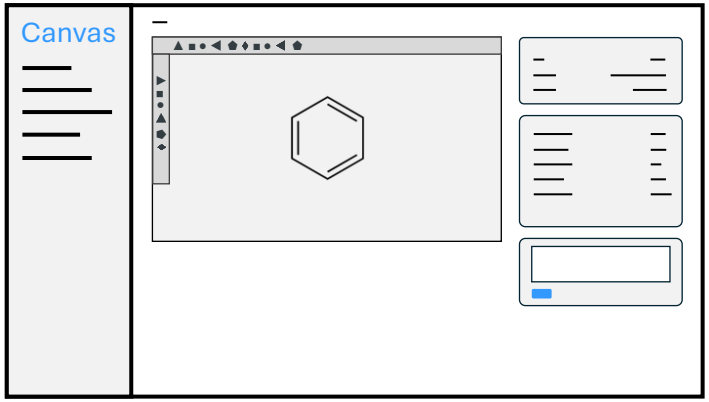
upload new file
or select recent files

Summary Page



select
compound

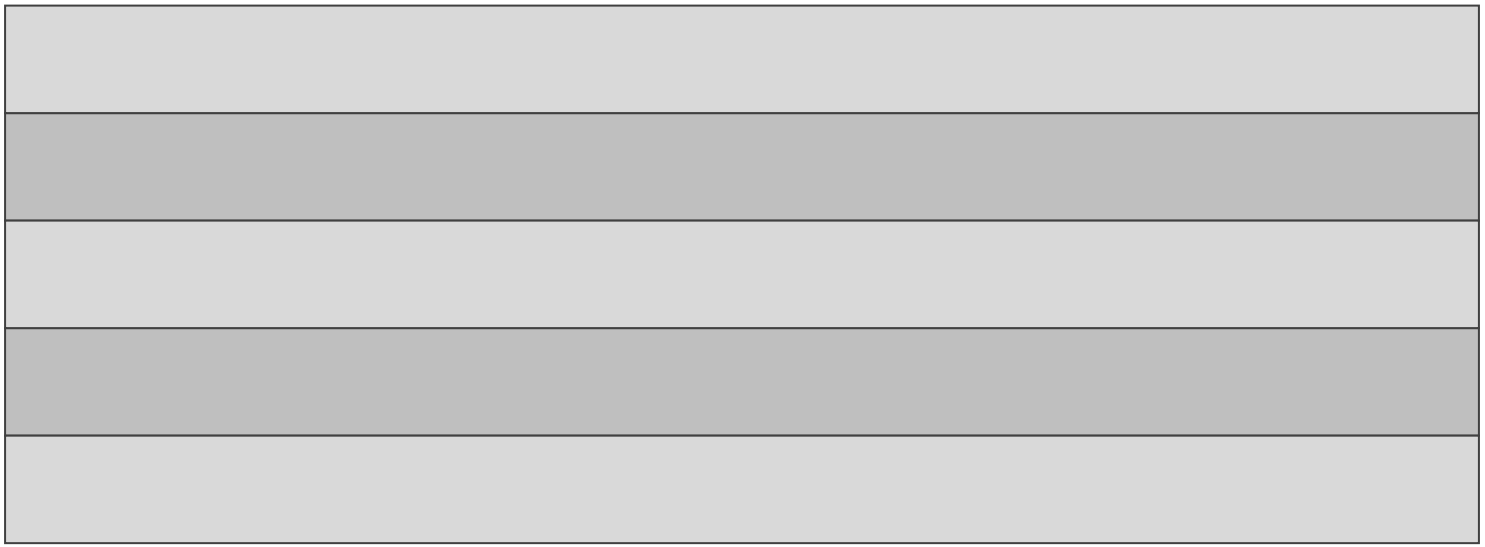
Annotation Page



Home Page

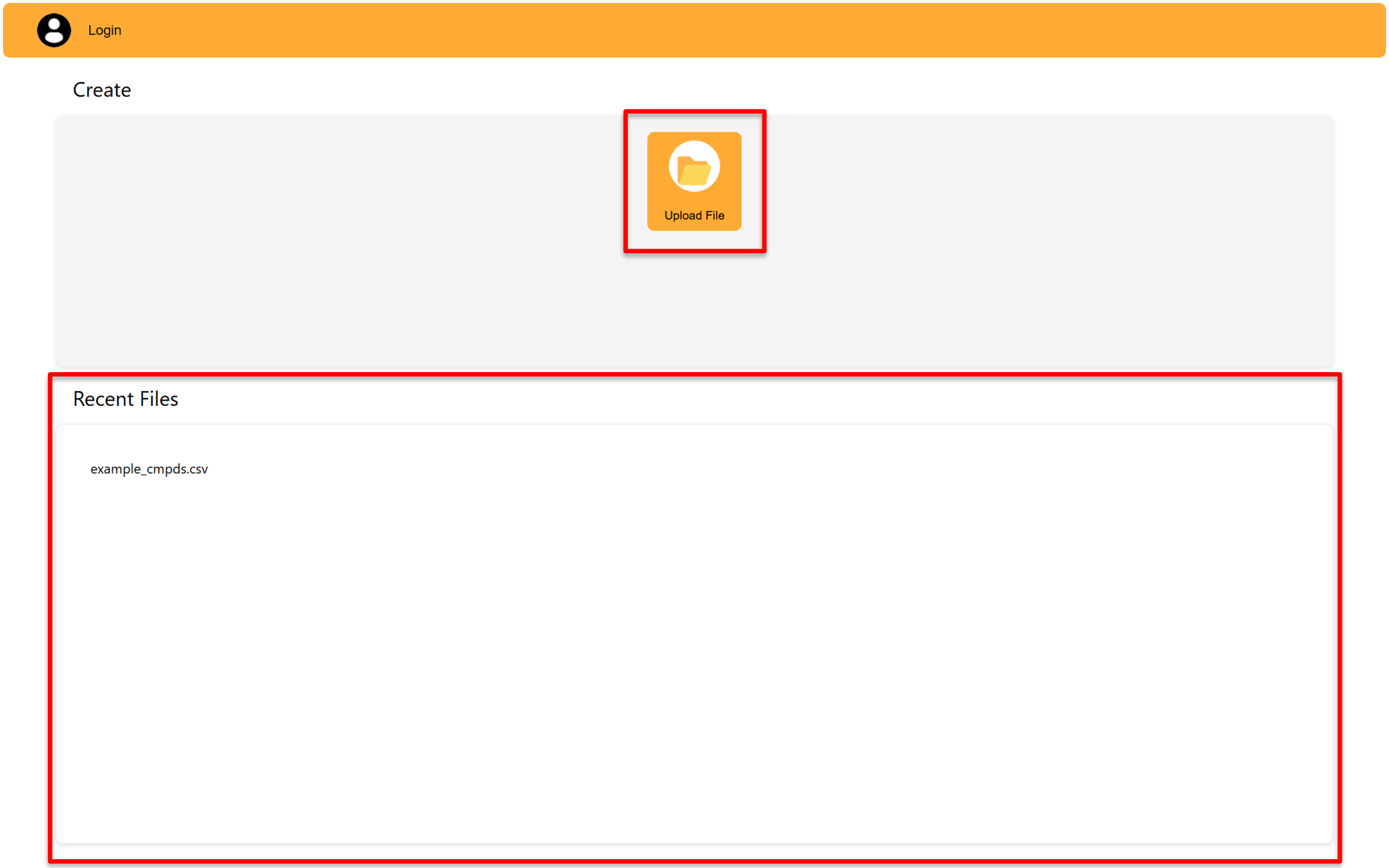
Create

Recent



Home Page

where users begin by uploading a dataset



Summary Page

where user can preview uploaded CSV file

Homepage

CSV Preview

cmpd_id	SMILES
cmpd_0	CCS(=O)(=O)c1ccccc1C(=O)N1CCO[C@H](C)C1
cmpd_1	O=C(OCC(=O)N1CCCC1)c1ccccc1OC(F)F
cmpd_2	Cn1cc[nH+]c1C[C@@H]1CCCN(C(=O)N[C@@H]2C[C@@H]3CCCc4cccc2c43)C1
cmpd_3	CCC(CC)(NC(=O)N(C)C1CC[NH2+]CC1)C(=O)[O-]
cmpd_4	CC(=O)[C@H](C)Sc1nc2sc3c(c2c(=O)n1C)CC[C@H](C)C3
cmpd_5	Cc1ccc(NC(=O)C2CCN(S(=O)(=O)c3cccs3)CC2)cc1Cl
cmpd_6	COC(=O)[C@]1(C#N)C(C)(C)[C@@]1(C#N)C(=O)OC
cmpd_7	Cc1ccc(Cl)cc1S(=O)(=O)N(Cc1ccsc1)Cc1ccco1
cmpd_8	C[NH2+]CC(C)(C)COc1ccccc1F
cmpd_9	O=C(COC(=O)c1ccc(-n2cnnn2)cc1)Nc1cc(Cl)ccc1F
cmpd_10	Cc1ccccc1OCCCN(C)C(=O)N[C@@H]1C=C[C@H](CO)C1
cmpd_11	Cn1c(CCNC(=O)c2cc(Br)cnc2Cl)n[nH]c1=S
cmpd_12	CCOC(=O)c1ccc2c(c1)nnn2C
cmpd_13	C[C@@H]1CC[C@@H](C(N)=O)C[NH+]1CCOc1ccccc1
cmpd_14	Cc1cccc(/C=C/C(=O)Nc2ccc3c(c2)OCCCO3)c1
cmpd_15	O=S(=O)(N[C@H](CC(F)(F)F)c1ccc(F)cc1)c1ccccc1F
cmpd_16	Cn1cc(/C=C/C(=O)c2cccc(F)c2)c(=O)n(C)c1=O
cmpd_17	Cc1nc([N+](=O)[O-])cn1CC(=O)Nc1ccccc1Cl
cmpd_18	C[C@@H](C(=O)Nc1cccc2ncccc12)[NH+]1CCC(C[NH+](C)C)CC1
cmpd_19	CC(C)[C@@H](CNc1ncc(C(N)=O)cc1Cl)N1CC[NH+](C)CC1

Annotation Page

where users engage directly with molecular structures, make modifications, and analyze data

Chemical Canvas

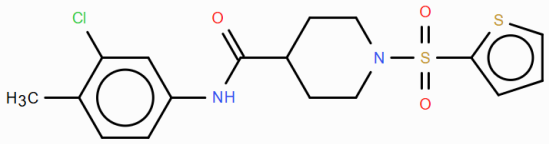
Current Mode: Annotate

Molecule: cmpd_5

- Modify
- Annotate
- Similarity Search
- Compute
- Export

Toolbar

Ketcher



Property Table

Basic Information

ID:	cmpd_5
SMILES:	Cc1ccc(NC(=O)C2CCN(S(=O)(=O)c3ccc s3)CC2)cc1Cl
Source:	example_cmpds.csv

Properties from CSV

binary_occ:	2.09
cont_occ:	0.52
low_gsh_prob:	0.26
med_gsh_prob:	0.60
high_gsh_prob:	0.14
selectivity:	0.02

Annotations

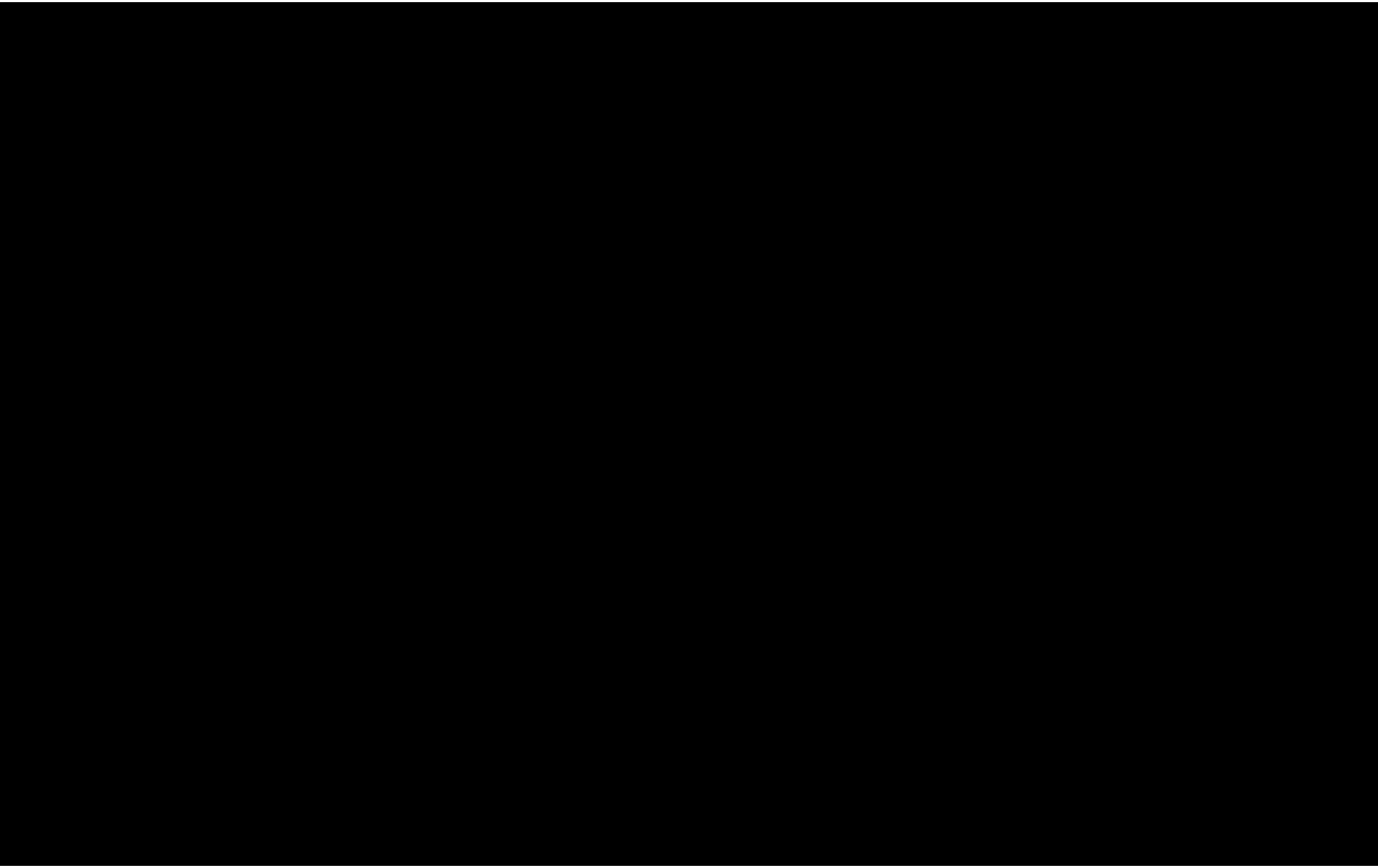
Add your annotations here...

Save Annotations

Annotation Box

← Back

Demo



Technologies

Technologies

Frontend

- Ketcher
  > a web-based chemical structure editor
  > Open-source and free to use
  > User Friendly
  > Easy to be embedded into React
- React
- Ant Design

Technologies

Backend and Libraries

- Flask
  > Simple and easy to learn
  > High flexibility
- RDkit
  > Molecular Representation and Handling
  > Molecular Similarity Comparison
  > Molecule Annotation Display
- Unittest
- torch_geometric(Spring Quarter)
  > Molecule Property Prediction

What we wished we knew earlier

Reflecting on ChemE 546

- Familiarization with GitHub – Actions, Projects, Forks, etc.
- Frontend design – wireframes + UI flow diagrams

Spring Quarter Plans

MVP and Beyond

- Close to meeting MVP goals!
  - Have individual functionalities – need to combine
  - Plan to finish by beginning of Spring Quarter
- Additional Functionalities
  - Separate summary page to display all uploaded molecules and respective annotation status
    - Exported as a new .csv file
  - Embed and display 3D structures of molecules
  - Embed and display protein data bank
  - Automate similarity search - compare with external database
    - Display most similar compounds
  - Automate property prediction

Thank you!! ☺☺☺☺

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

CHEMICAL ENGINEERING
UNIVERSITY of WASHINGTON

Flagship Pioneering