

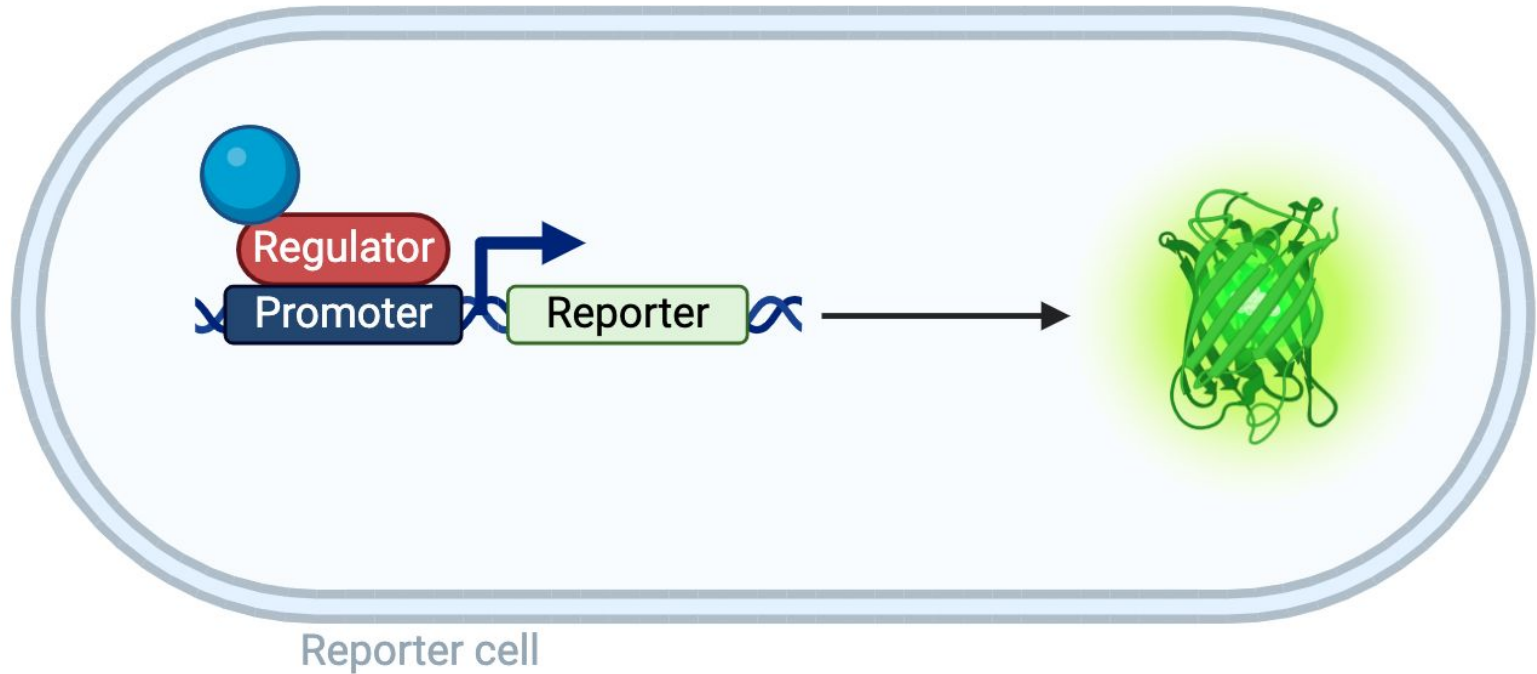


Ligify-Reverse

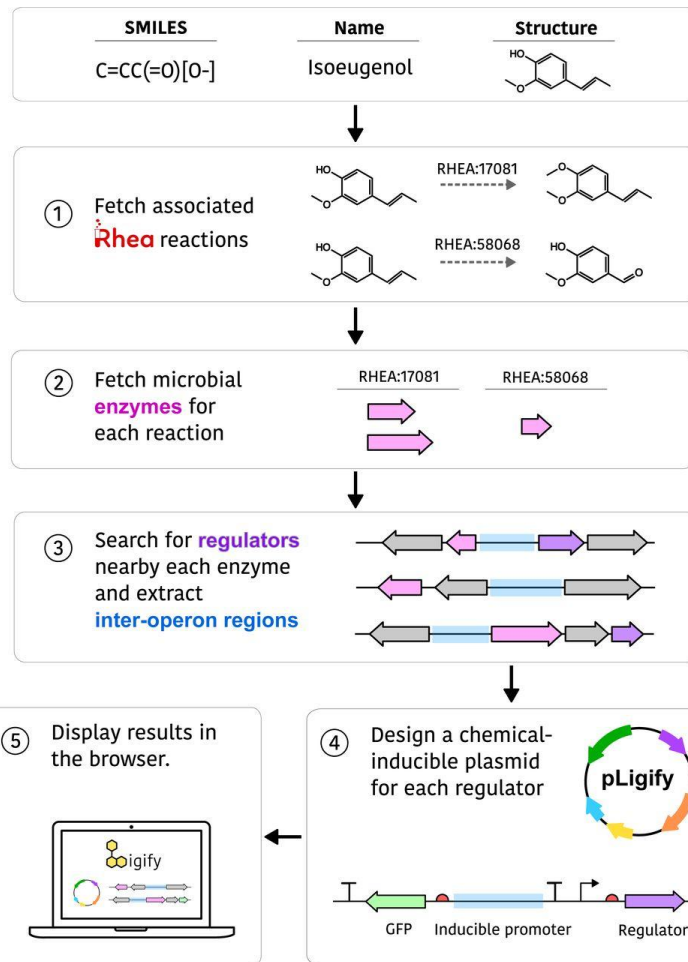
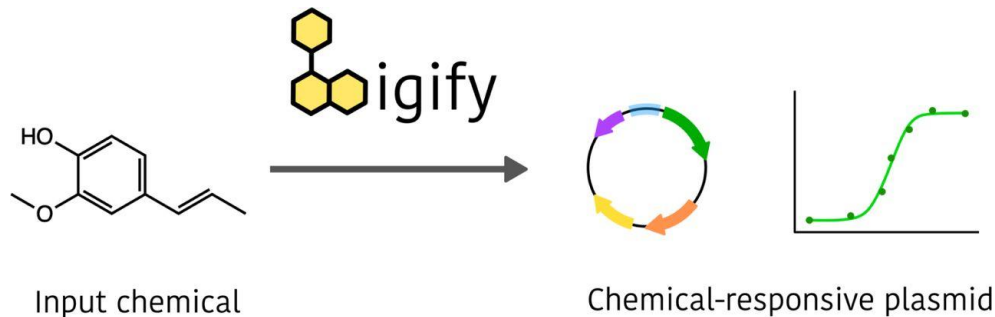
Silver Lab Rotation Project



Motivation / Problem



Ligify

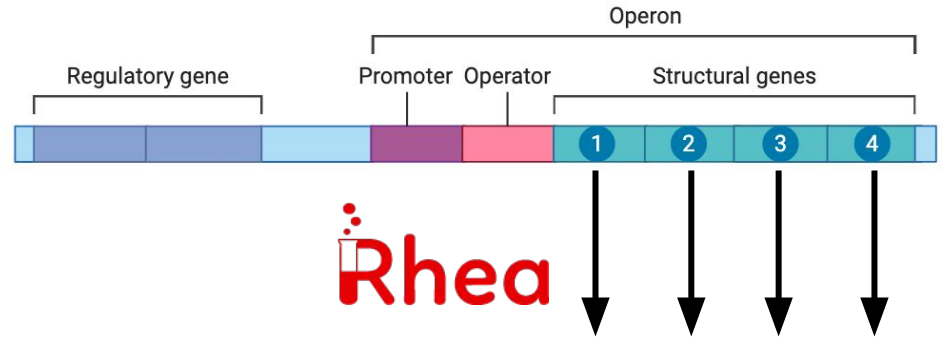


Ligify-Reverse

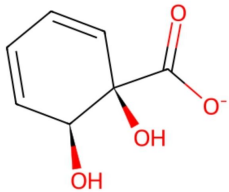


BLAST

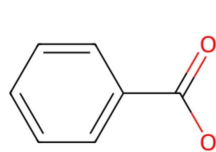
WP_212515502.1
WP_025094908.1
WP_151815971.1
WP_151840678.1
WP_288396828.1
AAC46441.1
WP_055414531.1
WP_088459563.1
WP_218262003.1
...



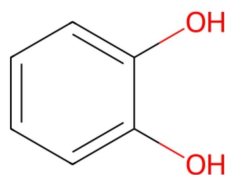
1st place



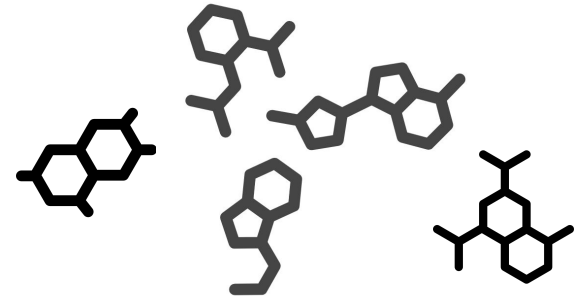
2nd place



3rd place



scoring



Adjustable Parameters

Run Mode

Would you like to search for homologs? If so, would you like to run using a DIAMOND database (faster) or remote BLAST (slower but more homologs)?

NO homolog search



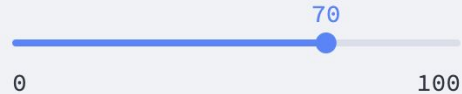
NO homolog search

Homolog search: DIAMOND dat...

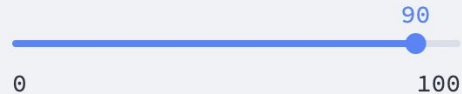
Homolog search: remote BLAST

BLAST Parameters (Ignore if not running homology search)

Identity Cutoff (%)



Coverage Cutoff (%)



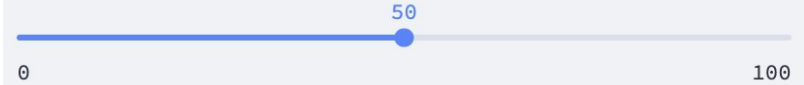
☒ Filter Redundant Homologs

Maximum Number of Homologs

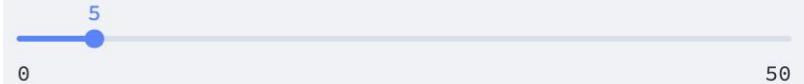
50

- +

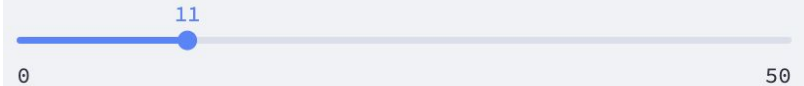
Score Cutoff



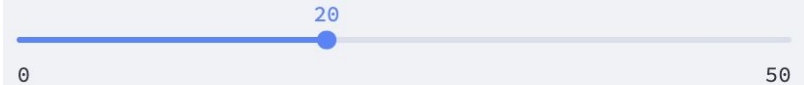
Weight: Distance From Regulator in Operon



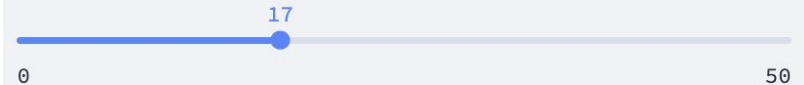
Weight: Identity (%)



Weight: Coverage (%)



Weight: Number of Appearances



Penalty for Missing in Query Protein's Operon



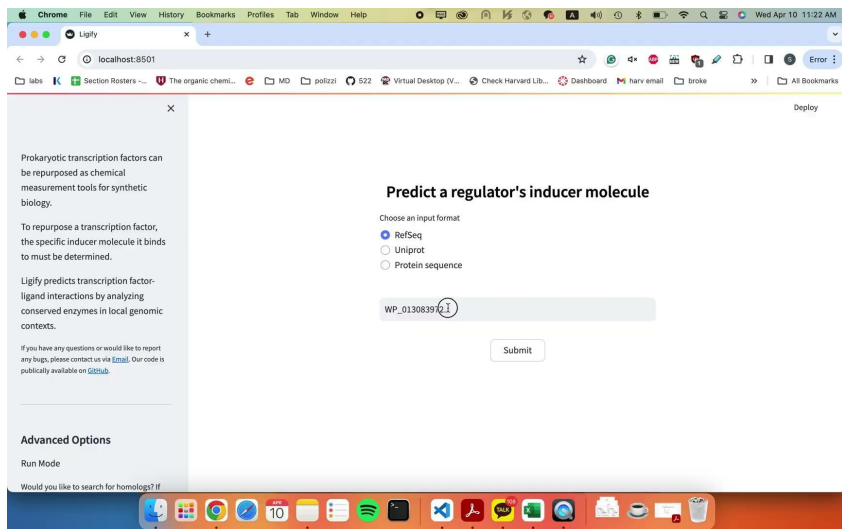


Demo Videos

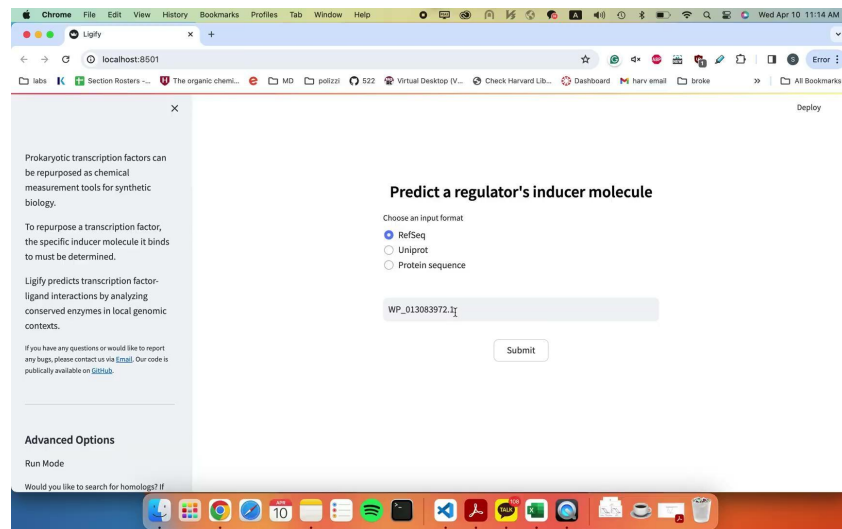


WP_011731512.1 (GylR) : Success

Known ligand: sn-glycerol-3-phosphate



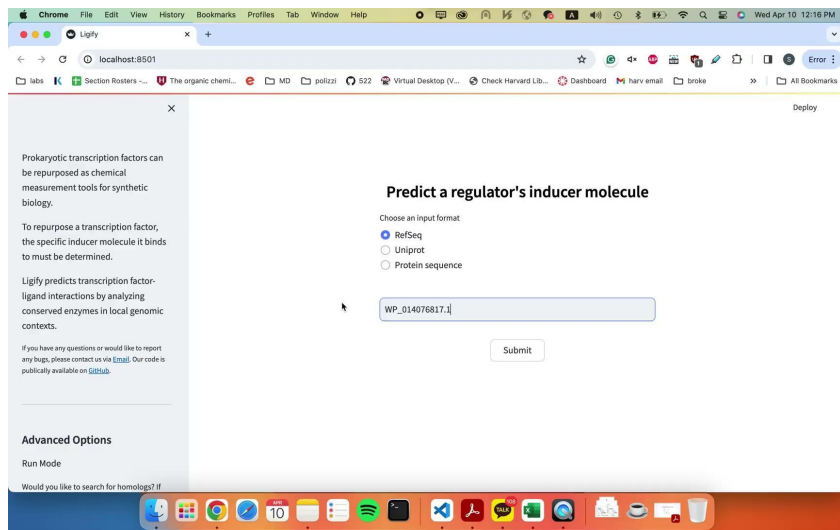
DIAMOND database



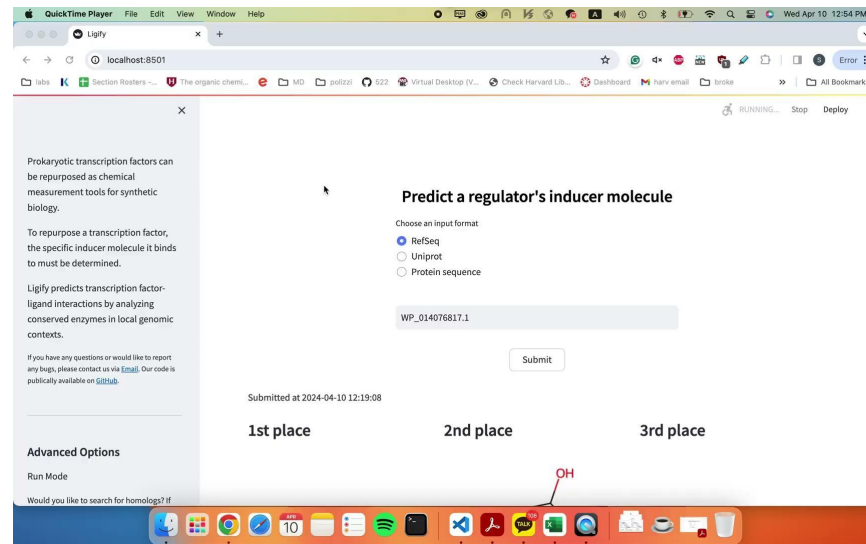
Remote BLAST

WP_014076817.1 (DesX): Success with remote BLAST

Known ligand: Syringate



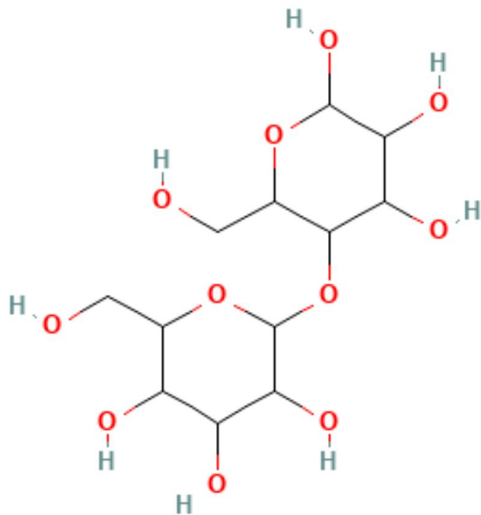
DIAMOND database



Remote BLAST

WP_011291385.1 (CelR) : Known ligand not found in list

Known ligand: Cellobiose



Run Mode

Would you like to search for homologs? If so, would you like to run using a DIAMOND database (faster) or remote BLAST (slower but more homologs)?

Homolog search: remote BLAST

BLAST Parameters (Ignore if not running homology search)

Identity Cutoff (%)
0 70 100

Coverage Cutoff (%)
0 90 100

☒ Filter Redundant Homologs
Maximum Number of Homologs
50 - +

Chemical Scoring

The ligands excluded from ranking are the following: L-S dihydrofolate, aryl GABA-L, arg-GABA-L, ADP-L, adenylylated, ATPase, AMP(2'), dATP(2') triphosphate, ADP(2'), carboxy acid anion, coenzyme A(2'), copper(2+), copper(2+)-dioxigen, disulfhydryl(2'), DNA phosphate

Predict a regulator's inducer molecule

Choose an input format

☒ RefSeq☐ Uniprot☐ Protein sequence

WP_011291385.1

Submit

Submitted at 2024-04-17 20:24:31

Input

Annotation: HTH-type transcriptional regulator CelR

Organism: Thermobifida fusca

Lineage: Bacteria, Actinomycetota, Actinomycetes, Streptosporangiales, Nocardioseae, Thermobifida

BLAST results

NCBI id	Identity	Coverage
---------	----------	----------

WP_011614884.1 (PhgR): No chemicals returned

Known ligand: Phenylglyoxylate

The screenshot shows a web browser window with the Ligyfy application running on localhost:8501. The application has a sidebar on the left with a description of its function: "Prokaryotic transcription factors can be repurposed as chemical measurement tools for synthetic biology. To repurpose a transcription factor, the specific inducer molecule it binds to must be determined. Ligyfy predicts transcription factor-ligand interactions by analyzing conserved enzymes in local genomic contexts." Below this is an "Advanced Options" section with a "Run Mode" dropdown set to "Homolog search: remote BLAST".

The main content area is titled "Predict a regulator's inducer molecule". It includes a "Choose an input format" section with radio buttons for "RefSeq" (selected), "Uniprot", and "Protein sequence". The input field contains the sequence identifier "WP_011614884.1". A "Submit" button is located below the input field.

Below the submission area, the status "Submitted at 2024-04-19 02:21:06" is displayed. The "Chemical scoring" section shows an "empty" result. The "BLAST results" section displays a table with the following data:

	NCBI Id	Identity	Coverage
1	U087123.1	97.826	100

The "Summary of all the enzymes found" section shows a table with the following data:

Enzyme	Frequency
0 benzoylformate decarboxylase	17

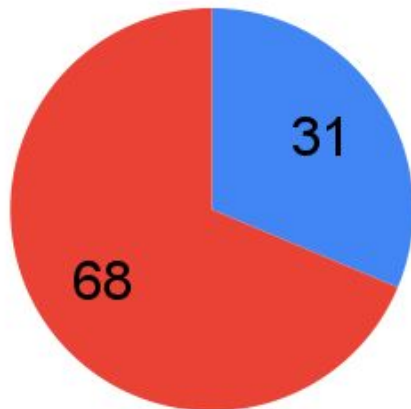
The bottom of the screenshot shows a macOS dock with various application icons.

Summary Statistics

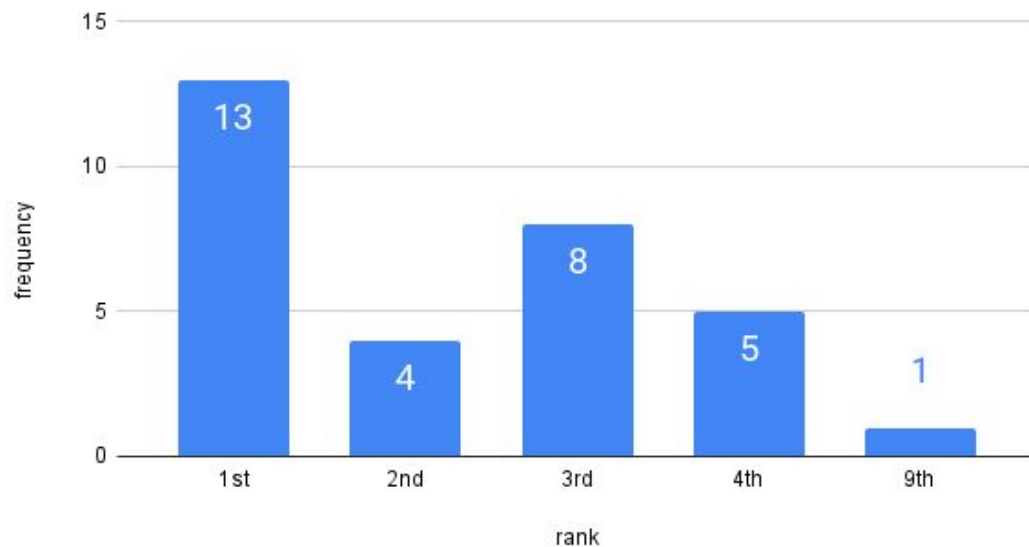
Benchmarked tool on 99 regulators with known ligands

Overall performance

- true ligand returned
- failures



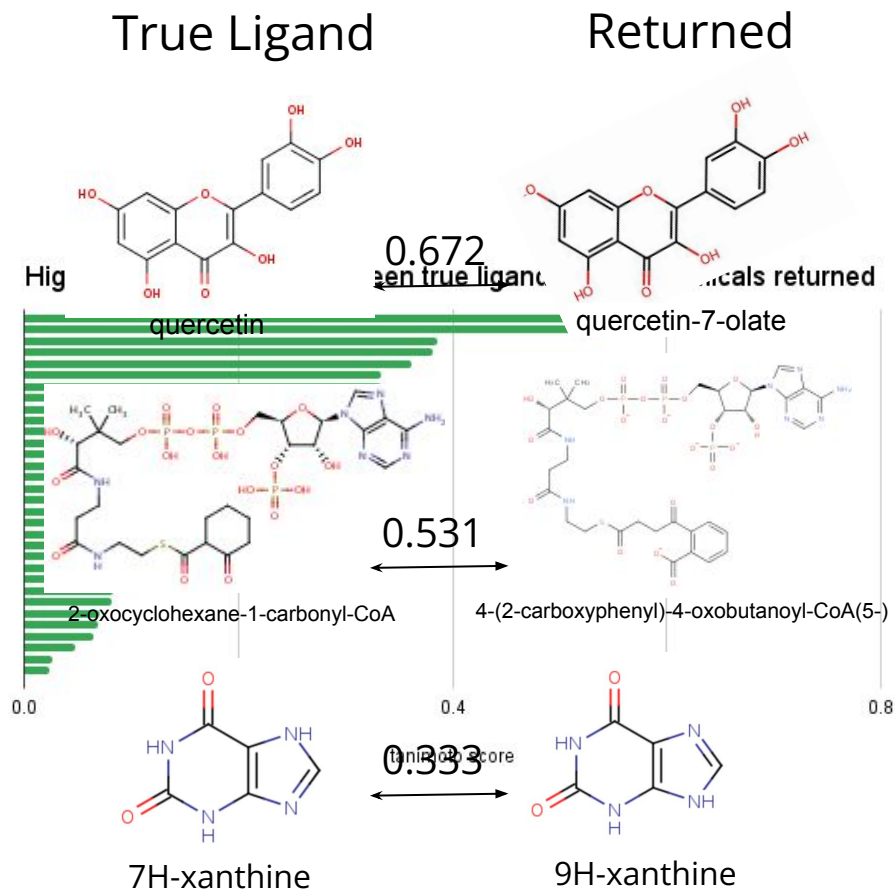
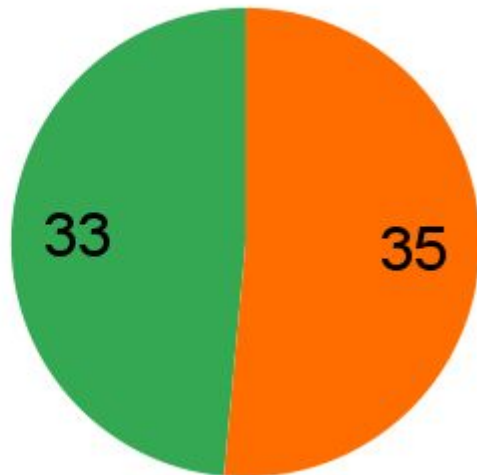
Success categorization: Rank of the true ligand



Summary Statistics

Failure categorization

- no chemicals returned from rhea search
- chemicals returned but not include target



Future Directions

- Query databases other than Rhea
 - ex) KEGG
- Continue to optimize chemical scoring function
- Using a similarity metric other than tanimoto to obtain more insight on the fails
- Apply tool to unknown transcription factors and test it out in lab

Supplementary

$$\text{score} = (1000 - \text{not present in query operon penalty} - (\text{distance avg} * \text{distance weight}) - (100 - \text{identity avg}) * \text{identity weight} - (100 - \text{coverage avg}) * \text{coverage weight} - (1 - \text{normalized occurrence}) * \text{num occurrence weight}) / 10$$

3/31 Ligify successes don't work in Ligify-reverse (NP_391277.1, WP_002965779.1, WP_003227022.1)

3/68 Ligify failures return the true ligand in Ligify-reverse (WP_010813655.1, WP_014076817.1, CAY46636.1)