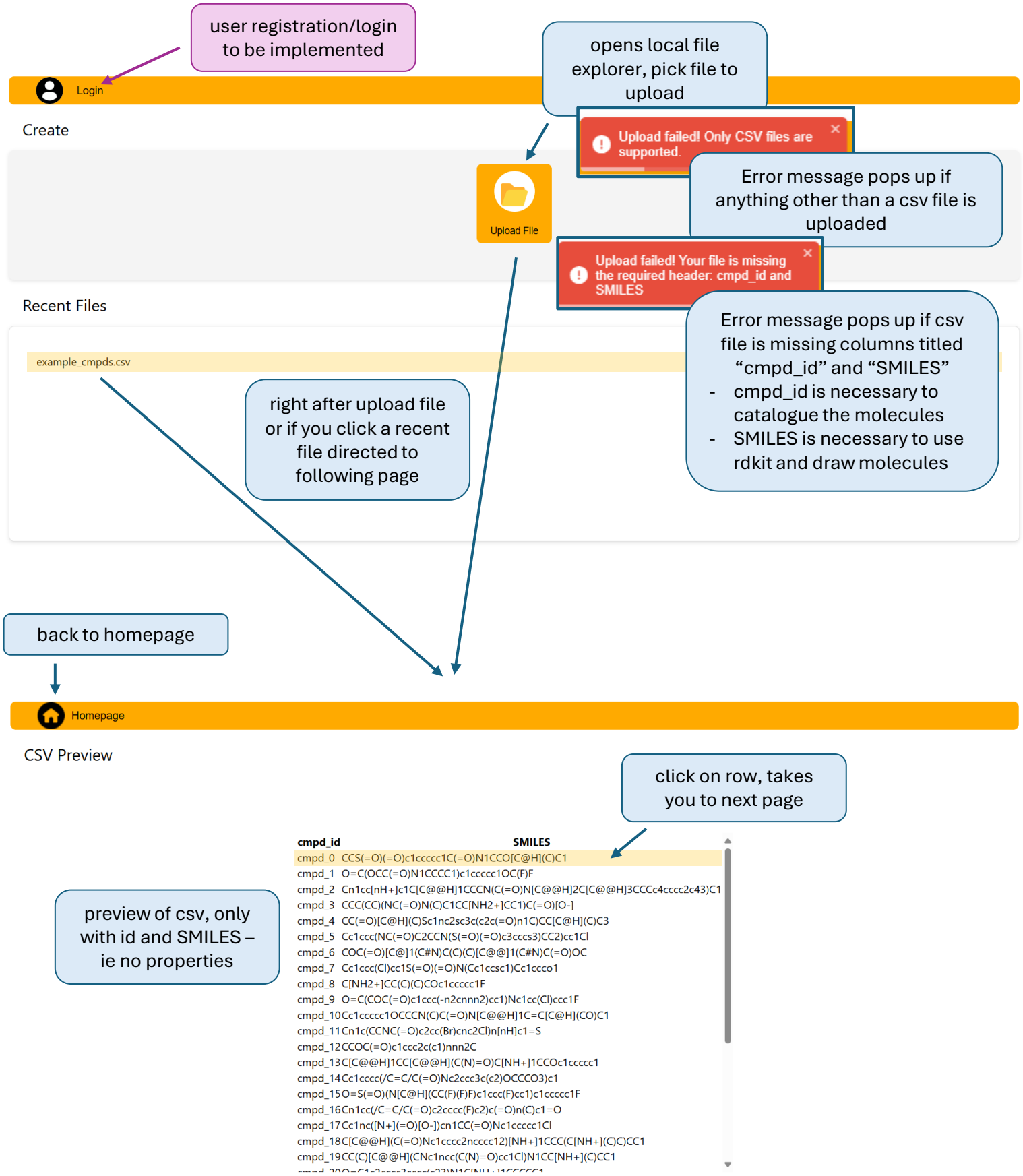
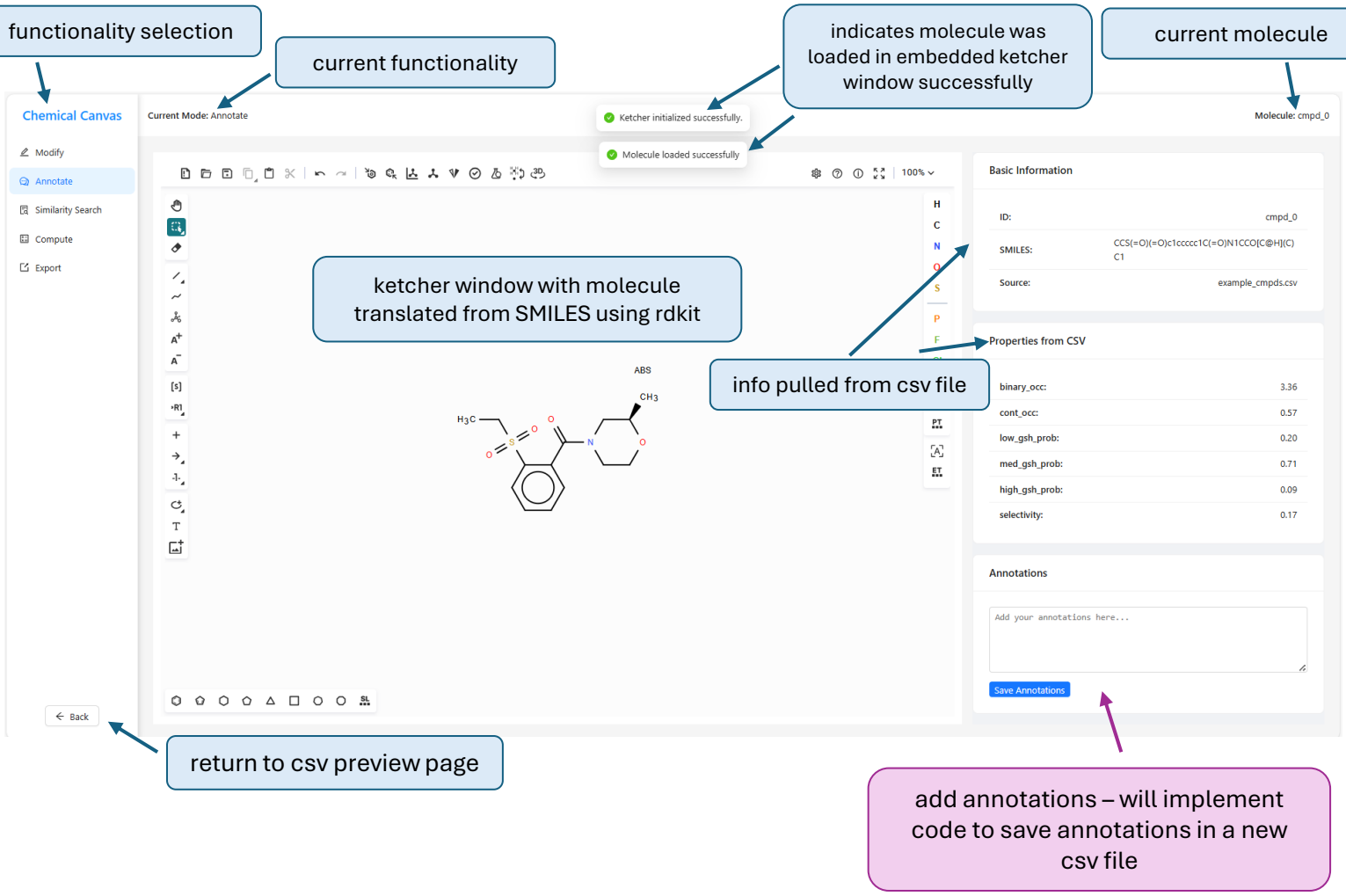
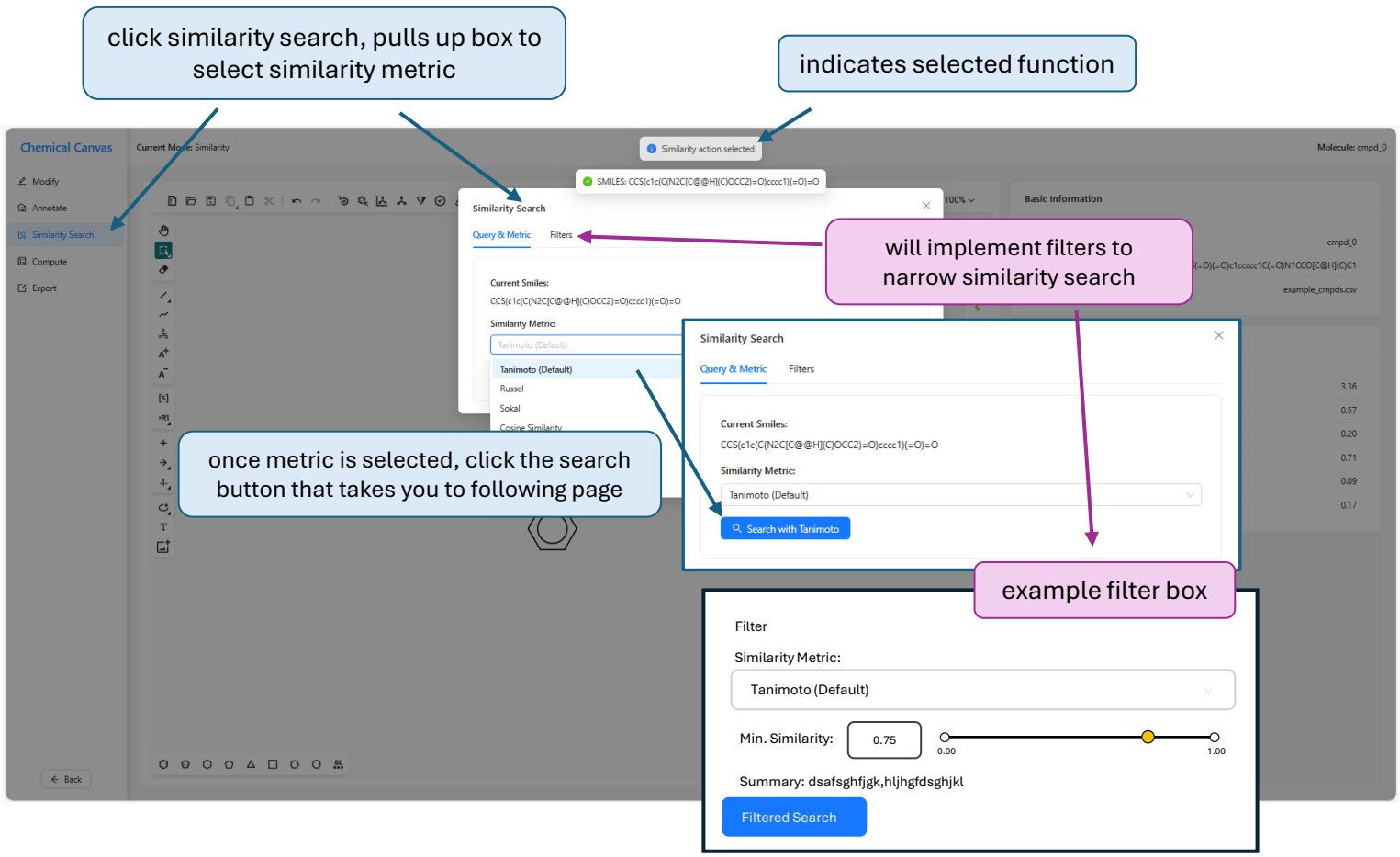
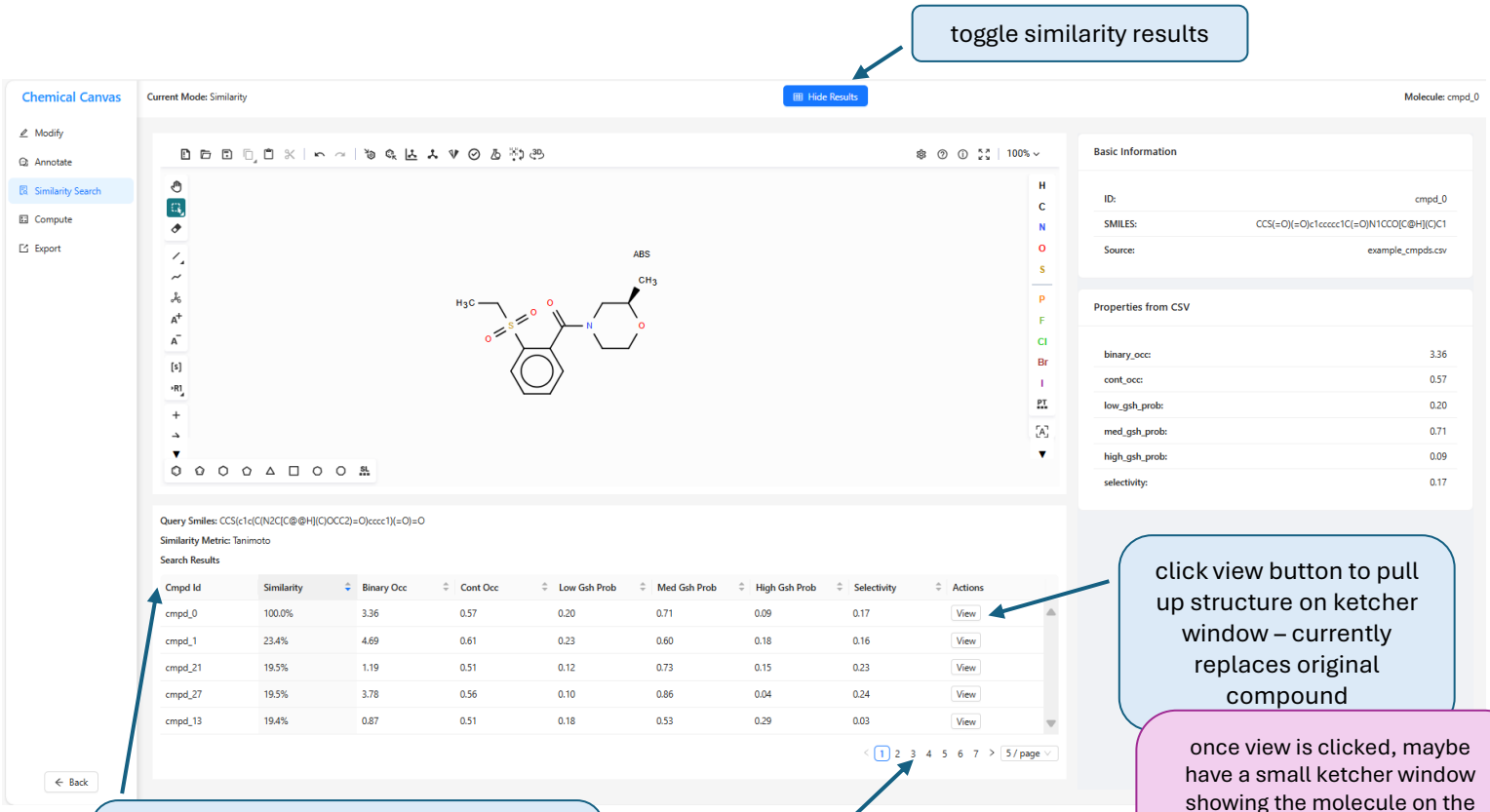

user registration/login to be implemented
Login
opens local file explorer, pick file to upload
Create
Upload failed! Only CSV files are supported.
Error message pops up if anything other than a csv file is uploaded
Upload File
Upload failed! Your file is missing the required header: cmpd_id and SMILES
Recent Files
Error message pops up if csv file is missing columns titled "cmpd_id" and "SMILES"
- cmpd_id is necessary to catalogue the molecules
- SMILES is necessary to use rdkit and draw molecules
example_cmpds.csv
right after upload file or if you click a recent file directed to following page
back to homepage
Homepage
CSV Preview
click on row, takes you to next page
preview of csv, only with id and SMILES – ie no properties
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


functionality selection
current functionality
indicates molecule was loaded in embedded ketcher window successfully
current molecule
Chemical Canvas
Modify
Annotate
Similarity Search
Compute
Export
Current Mode: Annotate
Ketcher initialized successfully.
Molecule loaded successfully
Molecule: cmpd_0
ketcher window with molecule translated from SMILES using rdkit
info pulled from csv file
Basic Information
ID: cmpd_0
SMILES: CCS(=O)(=O)c1ccccc1C(=O)N1CCO[C@H](C)C1
Source: example_cmpds.csv
Properties from CSV
binary_occ: 3.36
cont_occ: 0.57
low_gsh_prob: 0.20
med_gsh_prob: 0.71
high_gsh_prob: 0.09
selectivity: 0.17
Annotations
Add your annotations here...
Save Annotations
Back
return to csv preview page
add annotations – will implement code to save annotations in a new csv file


click similarity search, pulls up box to select similarity metric
indicates selected function
Chemical Canvas
Current Mode: Similarity
Similarity action selected
Molecule: cmpd_0
Similarity Search
Query & Metric
Filters
will implement filters to narrow similarity search
Current Smiles:
Similarity Metric:
Tanimoto (Default)
Russel
Sokal
Cosine Similarity
Search with Tanimoto
once metric is selected, click the search button that takes you to following page
example filter box
Filter
Similarity Metric:
Tanimoto (Default)
Min. Similarity:
0.75
0.00
1.00
Summary: dsafsghfjgk,hljhgfdsghjkl
Filtered Search


toggle similarity results
Chemical Canvas
Current Mode: Similarity
Hide Results
Molecule: cmpd_0
Modify
Annotate
Similarity Search
Compute
Export
Basic Information
ID: cmpd_0
SMILES: CCS(=O)(=O)c1ccccc1C(=O)N1CCO[C@H](C)C1
Source: example_cmpds.csv
Properties from CSV
binary_occ: 3.36
cont_occ: 0.57
low_gsh_prob: 0.20
med_gsh_prob: 0.71
high_gsh_prob: 0.09
selectivity: 0.17
Query Smiles: CCS(c1c(C(N2C[C@@H](C)OCC2)=O)cccc1)(=O)=O
Similarity Metric: Tanimoto
Search Results
Cmpd Id | Similarity | Binary Occ | Cont Occ | Low Gsh Prob | Med Gsh Prob | High Gsh Prob | Selectivity | Actions
cmpd_0 | 100.0% | 3.36 | 0.57 | 0.20 | 0.71 | 0.09 | 0.17 | View
cmpd_1 | 23.4% | 4.69 | 0.61 | 0.23 | 0.60 | 0.18 | 0.16 | View
cmpd_21 | 19.5% | 1.19 | 0.51 | 0.12 | 0.73 | 0.15 | 0.23 | View
cmpd_27 | 19.5% | 3.78 | 0.56 | 0.10 | 0.86 | 0.04 | 0.24 | View
cmpd_13 | 19.4% | 0.87 | 0.51 | 0.18 | 0.53 | 0.29 | 0.03 | View
5 / page
Back
click view button to pull up structure on ketcher window – currently replaces original compound
once view is clicked, maybe have a small ketcher window showing the molecule on the side, under “Properties from CSV” box
- Auto-highlight similar substructures
similarity search parses through your csv file and presents the most similar compounds with their respective properties in order of similarity
currently can go through and see similarity with all other molecules in csv – page through them here + change number of rows per page
once filters are implemented, only those that satisfy the conditions should appear
export functionality clicked
Chemical Canvas
Current Mode: Export
Molecule: cmpd_0
Modify
Annotate
Similarity Search
Compute
Export
Basic Information
ID: cmpd_0
SMILES: CCS(=O)(=O)c1ccccc1C(=O)N1CCO[C@H](C)C1
Source: example_cmpds.csv
Properties from CSV
binary_occ: 3.36
cont_occ: 0.57
low_gsh_prob: 0.20
med_gsh_prob: 0.71
high_gsh_prob: 0.09
selectivity: 0.17
Export Options
Get SMILES
Get Molfile
SMILES:
CCS(c1c(C(N2C[C@@H](C)OCC2)=O)cccc1)(=O)=O
Molfile:
currently can only generate SMILES and molfile as plain text
eventually have an option to export molecule + annotations to a new csv file
- also should add function to create actual .mol file
eventually have ML implemented to predict properties of molecules...

Other stuff

	CompoundID	SMILES	molecule_annotation	highlighted_atoms
1	* cmpd_0	:(=O)N1CCO[C@H](C)C1	example annotation	[5, 6, 7, 8, 9, 10]
2	cmpd_1	CCCC1)c1ccccc1OC(F)F		[]
3	cmpd_2	@H]3CCCc4cccc2c43)C1		[]
4	cmpd_3	CC[NH2+]CC1)C(=O)[O-]		[]
5	cmpd_4	(=O)n1C)CC[C@H](C)C3		[]
6	cmpd_5	))(=O)c3cccs3)CC2)cc1Cl		[]
7	cmpd_6	;)[C@@]1(C#N)C(=O)OC		[]
8	cmpd_7	:O)N(Cc1ccsc1)Cc1ccco1		[]
9	cmpd_8	+]CC(C)(C)COc1ccccc1F		[]
10	cmpd_9	nnn2)cc1)Nc1cc(Cl)ccc1F		[]
11	cmpd_10	)@H]1C=C[C@H](CO)C1		[]
12	* cmpd_11	2cc(Br)cnc2Cl)n[nH]c1=S	other example annotation	[7, 6]
13	cmpd_12	)C(=O)c1ccc2c(c1)nnn2C		[]
14	cmpd_13	O)C[NH+]1CCOc1ccccc1		[]
15	cmpd_14	Nc2ccc3c(c2)OCCCO3)c1		[]
16	cmpd_15	F)c1ccc(F)cc1)c1ccccc1F		[]
17	cmpd_16	cccc(F)c2)c(=O)n(C)c1=O		[]
18	* cmpd_17	])cn1CC(=O)Nc1ccccc1Cl	third example annotation	[]
19	cmpd_18	-]1CCC(C[NH+](C)C)CC1		[]
20	cmpd_19	cc1Cl)N1CC[NH+](C)CC1		[]
21	cmpd_20	c23)N1C[NH+]1CCCCC1		[]
22	cmpd_21	[C@H]2Cc2ccccc2)ccc1F		[]
23	cmpd_22	NCc3ccccc3)CCC2=O)c1		[]
24	cmpd_23	)(=O)=O)C2)CCC[NH2+]1		[]
25	cmpd_24	=O)CNc1nc2cc(Br)ccc2s1		[]
26	cmpd_25	Nc1ccc(OC)nc1)c1ccccc1		[]
27	cmpd_26	CC(=O)N1CCc2ccccc2C1		[]
28	cmpd_27	cc3)CC2)cc1[N+](=O)[O-]		[]
29	cmpd_28	2CCC(C)(C)[C@H]2O)C1		[]
30	cmpd_29	C(=O)N(C)C2)nc2ccccc21		[]
31	* cmpd_30	Nc3ccc(Br)c(C)c3)CC2)n1		[4, 5, 6]

Written code to store annotations + highlighted atoms in a csv file

- Highlighted atoms determined by canonical atom ordering
    - Atom indices as determined by rdkit
- Molecule annotation would come from annotation box (screen 3)
- Highlighted atoms pulled from highlight feature in ketcher
    - How does ketcher save highlights (if it even does)??
- Compound with annotations and/or highlights are indicated with an asterisk near the compound id
- If uploaded csv is title example_cmpds.csv, the output csv will be called example_cmpds_annotated.csv
    - i.e. <original file name>_annotated.csv
    - Table below is made using example_cmpds.csv file
- Need to implement to site
- Maybe this csv can be saved and automatically show up in recent files (screen 1)

Additional things to think about

- Saving stuff
  - Save ketcher edits
    - Save/save-as button for ketcher edits and annotations
  - Auto-save??
  - Once user registration implemented – save to user profile/workspace
- Implementing similarity search across multiple files
  - Multiple files the user uploads
  - External database?
- If we have large file (millions of molecules), how can we efficiently conduct similarity search
  - In the backend exclude molecules past a certain threshold of molecular weight, then perform similarity test on those remaining
    - Molecules of vastly different MWs are very unlikely to have similar properties
  - Add more filters for users to fine tune their searches
    - Molecule type
    - Molecular weight (as above)
    - Contains certain substructure
    - Boiling point/melting point/freezing point
    - pH
    - etc.
- Generate additional properties from external databases
- Implementing ML to predict molecule properties not found in external databases
- Embed and display 3D structures of molecules with their atomic properties mapped to atom nodes
- Embed and display PDB of protein target associated with a given protein-ligand interaction
- Option for running selected compounds through ML/computational backend pipelines