Small Molecule Search

David Danko and Noah Daniels

Small Molecule Search

Different Approaches

- Functional Groups
- Molecule Properties
- Structural Similarity

Structural Similarity

- Match molecules by the size of common subgraph
- Most general approach
- NP Hard

Structural Similarity

- fmcsR (Wang, Backman, Horan, Girke)
- SMSD (Rahmna, Bashton, Holliday, Schrader, Thornton)
- Approximate Solutions to MCS that run in polynomial time
- Can be used to do a single fold search

Our Approach

1. Compress Database into a set of representatives

2. Search for MCS between rep(query) and the set of representatives

3. Search for MCS between the query and the molecules corresponding to the results of step 2

Two fold approach

Compression

- Molecules can be represented as graphs (atoms as nodes, bonds as edges)
- Unlabeled, unweighted graph yielded 30-50% representatives to molecules
- Removing edges and nodes not part of a cycle yielded ~5% representatives to molecules

Talk about how the compressed database is actually made

Search

- Initial approach to search should have speedups proportional to compression ratio
- May be less precise than single fold
- Process of compression suggests second approach to search, likely accurate but will produce fewer results

Talk about the two different types of search

Sometimes the results of middling accuracy are most useful

Current Progress

- Compression is done
- Search is almost ready to start testing
- Initial results will be from the simple case of search
- Distributable tool will be ready for launch soon after search is tested