

Searching for reaction pathways in formose data

1. Generate reaction network:

Contains reactions for:

- enolate formation,
- enolate protonation,
- aldol addition of formaldehyde to enolates,
- Cannizzaro reactions and
- enolate reactions with sugars of $C_{\leq 4}$.

2. Remove nodes

Formaldehyde, hydroxide and water are ignored as nodes in pathways searches ($CaCl_2$ is not included in the framework).

3. For each data set:

- Make a list of detected compounds, include which compounds cannot be detected.
- This list is ordered by amplitude (highest amplitude first)
- Only reactions which include compounds in this list as reactants are included in the pathway search.
- Remove reactions in which inputs are products

4. Connect inputs (reactants) to products

Find the shortest pathways from the carbon input to the compound with the highest amplitude .

5A. [Version 1] Connect products to each other.

Not satisfactory, some products arise as products of their own reactions (e.g. glycolaldehyde)

- iterate down the list of detected amplitudes, adding the shortest paths between consecutive members of the list.

5B. [Version 2] Connect products to each other.

better

iterate down the amplitude list, find shortest pathways to every compound downstream of the node.

6. Check that all products have reactions leading to them

If they do not, cycle backwards from them in the sorted amplitude list until a connection is found.

7. Interpret

Considering this route with step 5B, simplest reaction pathway, according to the stated reaction rules, is within the discovered networks. However, there may be extraneous reaction pathways which 'short circuit' the reaction network.

It may be that the amplitude sorting is not required to afford similar results. It may be worth performing searches using randomised list of compounds as a search framework. However, if randomised lists give similar results, the outcome is likely due to the constraints applied (see above) during the network searches: the logic applied is sufficiently refining as to exclude almost all irrelevant reaction pathways. If amplitude ordering does not matter, then the power of the explanation logic is

highlighted- logic excludes so many possible reaction pathways, that the level of detail provided by the amplitude information is degenerate.