### Imperial College London

# Updating the database of POCs

Andrew Tarzia

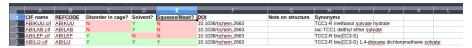
# Imperial College London Search process

- Extract all REFCODEs associated with list of 23 authors using
  ConQuest OR CSD Python API
- Filter by:
  - Has 3D coordinates
  - Not organometallic
  - Crystal structures only
  - Published post 2015
- Visualize in ConQuest and keep only 'cage-like' structures
- Produces 108 cage structures (CDB41 has 41 structures) although this may not be exhaustive

#### Imperial College London

# Properties of crystal structures

- Visualize all structures and determine if:
  - Cage backbone is disordered
  - Solvent is present
  - Squeeze/Mask is used



# Imperial College London Next steps

- Automatically clean up disorder
- Automatically remove solvent and extract cage molecules (pyWindow)
- The above process is also being applied to collect metal-organic cages from the CSD