Research Abbreviated Summary

* Compiled list of 335 bound antigen-antibody complexes from the researcher’s database
  + Isolated antigen chains with corresponding annotated residues
    - Limited complex (“bound”) number to 275 to ensure annotated residues are only exist on a single antigen chain per complex
  + Found corresponding researcher-RCSB chain letter identifiers
* From the list of 275, 195 unbound antigens were found in RCSB that had >95% similarity to the antigens in the complex.
  + Mapped annotated residues from bound antigen chain onto corresponding unbound antigen chain
    - Generated list of 195 unbound antigens with annotated data
  + Further modified about 35 proteins from the set of 195 to remove extraneous protein segments
* Following alignments and curation of bound and unbound sets, we began predictor calculations
* Currently have:
  + VORFFIP, METAPPIPSP, ISPRED, SPPIDER, DOCKPRED, DIscoTope(?)
    - For both bound/unbound
    - ISPRED gave us dynamic cutoff data
    - Will need to review this data to ensure completion
  + ISPIP
    - Completed for the bound
    - Unbound will require more analysis/work (ex: pymol)
* Clustering
  + Scripts developed
  + Capability to incorporate pymol
  + Have reviewed both pre-clustering and post-clustering F-score and MCC scores for the epitope predictions
* Planned future steps:
  + Review and fix up bound/unbound data
  + Review ISPIP unbound data
  + Better coordination through github