**PDB Projects Journal**

**Load PDB Structure(s) into MySQL:**

1. Download PDB from RCSB – Step 2 to batch download from a list of PDB IDs
2. Paste PDB IDs into <http://www.rcsb.org/pdb/download/download.do>
   1. Check PDB Format
   2. Select uncompressed
   3. Download all files
3. Feed PDB file to pdb\_to\_tranpep.py
   1. For multiple files, call *load\_pdb.sh* in a directory with all of (and only) those PDB files you would like to load.

**Identify Human Homologues from (Non-)Human Proteins in PDB Files:**

1. Query the UniGene Accession number for the UniProt ID
2. Test\_homology.pl is a test script for pulling the human homologue transcript for a UniGene ID. That transcript should then be able to pick up at the transcript conversion step in protein\_to\_genomic.pl, but the functionality isn’t finished. The code needs to be written, but the biggest hiccup in the pipeline is the UniGene lookup, which does not have an obvious way to automate yet. It might be easiest to download *all* conversions from UniProt->UniGene.

**Generate PDB Annotation File from MySQL:**

1. Open the MySQL database containing PDB information
2. Select columns: *pdbid,chain,seqres,x,y,z*
3. Save to space delimited file
4. In RVCLUST
   1. Load *rvclustobject* with PEDMAP information
   2. *rvobj <- annotate(rvobj,”PDB”,c(“PDB\_x”,”PDB\_y”,”PDB\_z”))*
   3. May also choose to use *PDBID, CHAIN*, or *SEQRES* for labeling or clustering

**Compute SDP Scores and Load Candidate PDB Structures:**

1. Feed FASTA file to group\_sim\_sdp.py
   1. This generates the SDP scores
2. Feed the SDP score output to fasta\_to\_pdb.py
   1. This loads the SDP scores into the database
3. Feed FASTA file to pull\_unp.py
   1. Pulls all uniprot IDs in the FASTA file
4. Load all relevant PDB structures into MySQL
   1. PDB\_query\_by\_unp.xml has a REST template for pulling PDB IDs by UNP
   2. Manually: <http://www.rcsb.org/pdb/software/wsreport.do>
   3. Automatically, learn how to script an “XML Post”
   4. Use this list to load the relevant PDB structures

**Mapping to PDB Structures with PyMol**

**Generate a PDB Cluster Map from RVCLUST**

1. Run the cluster analysis
2. Write the variants table to tab-delimited file with columns (in order): *pdbid,chain,seqres,value*
3. Open PyMol and fetch pdbid
4. In the cmd bar: *run overwrite\_pymol.py*
5. In the cmd bar: *overwrite\_bfactors(‘pdbid’,’tab-delim from step 3’)*

**Generate a PDB SDP Map from SDP Scores in MySQL:**

1. Open the MySQL database containing PDB information and SDP values
2. Join as necessary and select columns (in order): *pdbid,chain,seqres,value*
3. Save to tab delimited file
4. Open PyMol and fetch pdbid
5. In the cmd bar: *run overwrite\_pymol.py*
6. In the cmd bar: *overwrite\_bfactors(‘pdbid’,’tab-delim from step 3’)*