

1

Methods

1.1 Datasets

The primary dataset of heme-containing proteins (HEM) was composed by searching for proteins containing “HEM” in the PDB. 58 PDBs were kept after culling a much larger list. Not all PDBs were suitable for the investigation - some contain two ligands in one pocket, or two proteins per PDB (IDK HOw to say this properly rn)

A similar culling was done for the other ligands. From this work we ended up with the following dataset sizes: HEM 58, HEC ...

Notes for each dataset.