

```

      .
      .
      .
MOLECULE:  3N69_B      K=2      1.0
clst:    1ZZT_A  3N67_A  3N68_A  3N69_A
rep:     3N67_A
clst:    1ZZT_B  3N67_B  3N68_B  3N69_B  3N6A_A
rep:     3N67_B

```

```

MOLECULE:  3HXU_A      K=2      1.0
clst:    3HXU_A  3HXV_A  3HXW_A
rep:     3HXV_A
clst:    3HXX_A  3HXY_A
rep:     3HXX_A

```

```

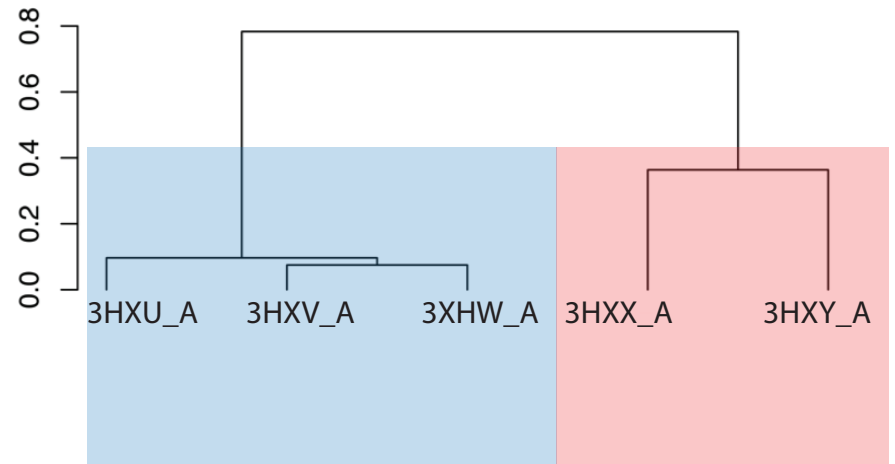
MOLECULE:  4KWW_D      K=2      0.979
clst:    4KVV_A  4KVV_B  4KVV_C  4KVV_D  4KVV_E
rep:     4KVV_A
clst:    4KWW_A  4KWW_B  4KWW_C  4KWW_D  4KWW_F
rep:     4KWW_B

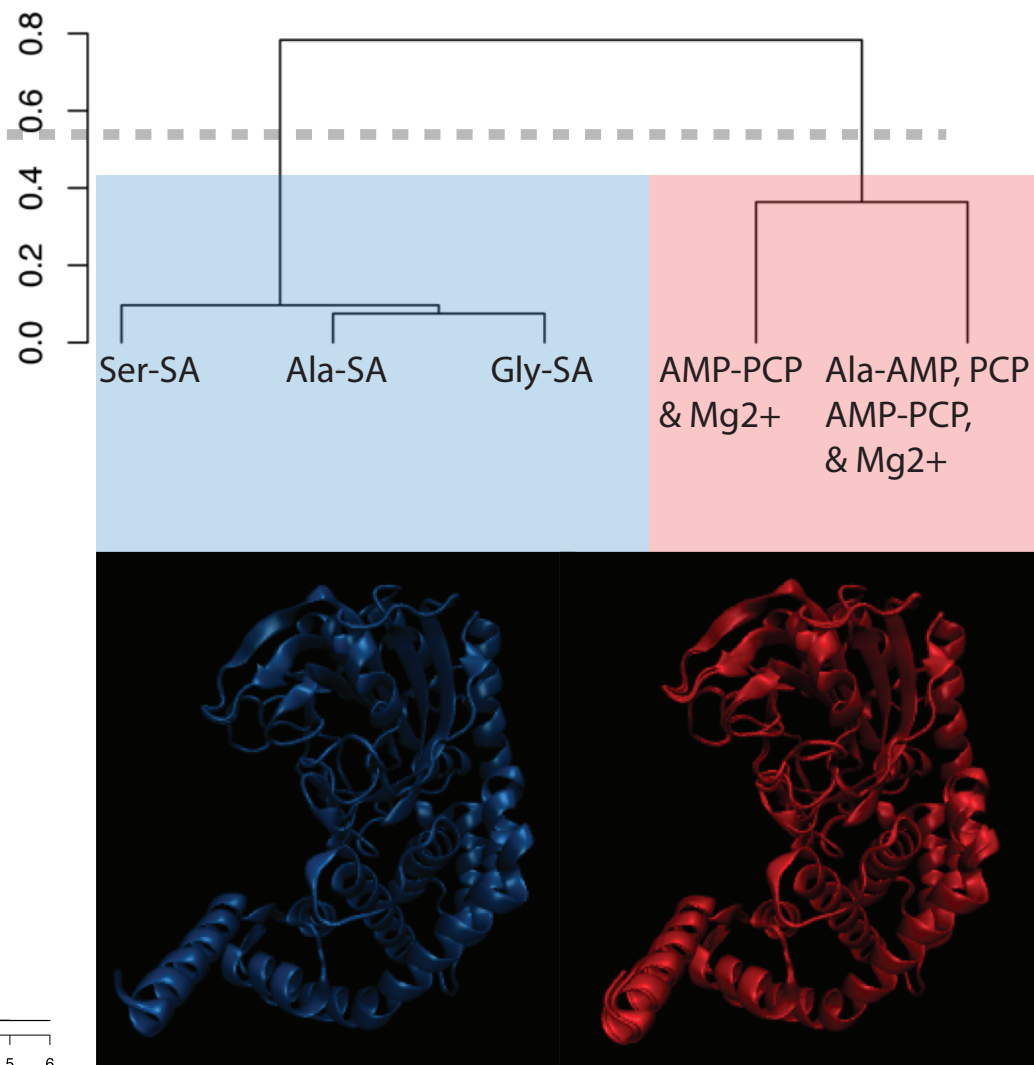
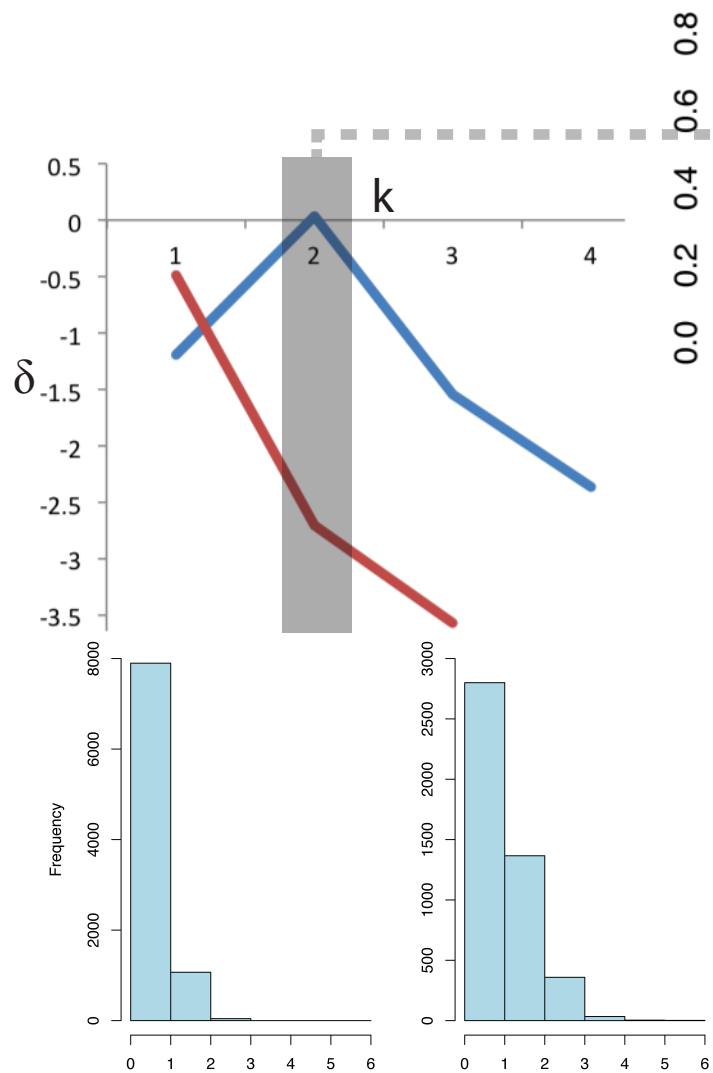
```

```

      .
      .
      .

```





Confounding factors:

Many cases in which the different clusters are not formed by diff PDBs, but rather by diff CHAINS. Ex:

clst 1: 1BD3_A 1BD3_C 1BD4_A 1BD4_C

clst 2: 1BD3_B 1BD3_D 1BD4_B 1BD4_D

Cases in which multiple chains from one PDB are the elements that form the entire sequence group

More common: cases in which different clusters do not really differ biologically (manual annotation).

Overlap with ASD:

num_eligible_predicted_allo_pdbs: 976

num_matches_to_ASAD: 131

total num entries in PDB: 82114 (as of 2012-06-15)

total num ASD PDBs: 9611

Fract of asd in pdb: 11.7%

num_matches_to_ASAD: 131

num_EXPECTED_matches_to_ASAD: 114 = (9611/82114) * 976