

Progress Report(Up until 30th December, 2015)

Anirudh Tiwari

Work Done

1. I was trying to reconstruct the data set as per Holland's paper. I was partially successful and created a data set of my own.

Results

The data set was constructed in the following steps:

1. All the chains which had the same number of domains as per CATH & SCOP were selected. I obtained around 80,000 chains from it.
2. I classified the obtained chains into Topological classes based on their Class, Architecture and Topology(as per CATH). I obtained 844 topological classes.
3. Now, for each of the classes, I tried to obtain a unique k-domain protein. This was done for every k from 1 to 5.
4. Following are the results

	Contiguous	Non-Contiguous	Total
1 Domain chains	677	-	677
2 Domain chains	118	29	147
3 Domain chains	38	23	61
4 Domain chains	10	5	15
5 Domain chains	6	8	14
			914

Holland's paper had a total of 355 chains, 106 1 domain, 138 two domain, 55 three domain, 8 four domain and 5 five domain chains.

Thus, it can be seen that my results differ greatly from Holland's paper, although I'll be starting with data set itself to test my methods to find the number of domains.

Next Steps

1. Test my domain finding methods on the constructed data set and hence, improve the methods.
2. Side by side, try to reconstruct the Holland's data set, so that we have a standard benchmark to test against.