Abstract

Enzymes are known to be remarkably specific to substrate/reaction they catalyze. However, many enzymes harbor inherent ability to catalyze non-native reaction/s or substrate/s that is referred as enzyme promiscuity. Such secondary reactions act as repertoire of novel catalytic activities, which could play an essential role in conferring fitness to an organism under genetic or environmental changes. Understanding the mechanistic basis of enzyme promiscuity can provide insights into enzyme evolution as well as facilitate rational design of desired substrate. In my thesis work, the aim is to explore general structural basis of promiscuity in enzymes. Further, as a case study, we performed structural analysis of *Escherichia coli* γ-glutamyl cysteine ligase (EcGCL) to gain insights into its substrate promiscuity.

In order to characterize general features of promiscuous enzymes, we systematically compared structural properties of active sites between *E.coli* generalists (promiscuous) and specialists (non-promiscuous) enzymes. Among various features compared, hydrophobicity and non-polar accessible surface area of binding pockets are predominant in promiscuous enzymes. Next, we compared ligand-induced local conformational changes in these two categories of enzymes that showed catalytic residues of generalists undergo relatively little (usually no) conformational change compared to specialists. This suggests that pre-organization of catalytic residues in generalists in aid the catalysis of alternate substrate s/reactions.

The catalytic residues of many enzymes are not yet characterized and most catalytic residue prediction methods do not provide their reliable ranked positions. To address this, we developed a consensus based catalytic residue prediction method (CSmetaPred). Here, normalized scores from four well-known predictors are combined as meta-score (average) to rank residues. Further including predicted pockets information with meta-score (CSmetaPred_poc) improved the prediction performance. The benchmarking results showed CSmetaPred_poc as the best performing method among methods assessed in this study. Importantly, CSmetaPred_poc improved ranked position of known catalytic residues and correctly identifies all catalytic residues within top 20 ranks for ~73 % of proteins in CSAMAC benchmark dataset. Both servers can be publicly assessed at http://14.139.227.206/csmetapred/.

To understand the structural basis of cysteine substrate promiscuity in EcGCL, we used docking and molecular dynamics simulation approaches. These analyses showed that substrate promiscuity is mostly governed by hydrophobic interactions and size of cysteine binding site. We classified three sequence diverged groups of GCL into subfamilies and performed function annotation of GCL identified members in completely sequenced genomes. Further, we characterized all substrate binding motifs of GCL families/subfamilies. The phylogenetic study of GCL showed members of group 2 and 3 are more closely related than group 1.