Inferred relationships by the combination of sequence, structure, and functional similarities can help understand the evolution of proteins (Doolittle, 1981; Grishin, 2001a; Murzin, 1998). In this thesis, we have attempted to classify proteins of the ferredoxin-like with the following objectives.

Objective 1 was to provide the structural classification of ferredoxin-like fold proteins based on inferred (using sequence/structure/function similarities) evolutionary relationships. We have identified and classified all the domains from PDB structures having at least a ferredoxin-like structure pattern. The overall framework of classification includes four levels: fold-groups (topology level), superfamilies (possible homology level), families (statistically significant sequence similarity/true homology) and domains. We have classified a total of 18996 FLDs into 155 Superfamilies and 395 Families. These families and superfamilies could be grouped into three distinguishable fold-groups named as ferredoxin-like CPI/IV, ferredoxin-like CPII/V, and ferredoxin-like CPIII/VI. Various circular permutation in FLDs was identified and added to the classification. The best example of three naturally permutated versions of FLDs (CPI/IV_vs_ CPII/V_vs_ CPIII/VI) was seen in three enzyme families LuxS, MPP, and ThrRS/AlaRS, where all the three ferredoxin CP types are evolutionarily related by structural similarity and functional residue conservations. Another example of inter-superfamily circular permutation (CPI/IV_vs_CPIII/VI) was observed in DN-DC domains of RND pump family and Rpb1_7 domain of RNA polymerase family members.

Intra-superfamily examples of naturally permuted versions of ferredoxin-like fold domains were also observed in 4Fe-4S ferredoxin-like (CPI/IV_vs_CPII/V), CcmK-like (CPI/IV_vs_CPIII/VI), Bacterial exopeptidase dimerization domains (CPI/IV_vs_CPIII/VI) and Origin of replication binding domain-like family members (CPI/IV_vs_CPIII/VI).

Our classification was also compared with widely used classification schemes such as ECOD, SCOP and CATH and our scheme group together the highest number of FLDs (18996) and superfamilies (155). The manual classification and comparative analysis of FLDs have led to the discovery of several novel FLDs which were previously classified as new folds by other protein databases. The SCFL also includes six novel FLD superfamilies to the classification *viz.*, MESD, LigT-like, the C-terminal domain of SAMHD1, Probable bacterial effector binding domains, FL insert domain in Prim-pol and N0 domain of secretin HofQ.

Objective 2 was to understand the sequence-structure-function relationships of the ferredoxin α+β barrel superfamily. The structural definition of ferredoxin α+β barrel includes two identical/non-identical FLDs that dimerize back-to-back to form a closed antiparallel eight-stranded barrel at the center and with helices positioned at both sides. The ferredoxin α+β barrel superfamily is one of the diverse superfamilies of FL fold where the closed α+β barrel structure has been studied as the basic functional unit for most proteins, while some families of this superfamily have been observed to adopt a duplicated and fused α+β barrel architecture, where two FLDs are connected by a flexible linker polypeptide.

The study identified a total of 1219 FLDs containing proteins in PDB (before 30 Dec. 2015) that form a complete ferredoxin α+β barrel architecture and grouped into 20 different ferredoxin α+β barrel families. For sequence conservation among diverged families, we prepared and analyzed a manual structure-based multiple sequence alignment (MSA) containing the representative domains of structurally characterized FL barrel proteins. The final MSA of ferredoxin α+β barrel superfamily representatives revealed two novel sequence motifs "FABB-1" (10 residues long) and FABB-2 (4 residues long) conserved in most

families. Careful analysis of "FABB-1" and "FABB-2" revealed four highly conserved aromatic side chain residues forming the hydrophobic core of the FLD cleft region. The observed novel sequence motifs along with the histidine conservations in FLDs can also be used for the functional prediction of hypothetical ferredoxin α+β barrel protein(s). The ligand mapping over the FLD containing barrels suggested the five major functional sites to the α+β barrel superfamily; two distinct sites at the ferredoxin-like cleft regions (in between the helix and sheet layer), one site between the loop 1 and loop 3, and two distinct sites at the opposite ends of the central barrel cavity.

The manual structure superimpositions along with the structure-based MSA between different heme-binding FLD families explain the plausible evolutionary mechanisms by which various heme-binding sites and barrels would have emerged (Acharya et al., 2016b). The study also finds the FLDs with two distinguishable modes of ferredoxin α+β barrel packing; type-1/IsdG-like and type-2/Yqjz-like, which differ in the orientation of the constituent domains. By showing examples with significant sequence and structure similarities between two types of barrel packing family domains, we explained the selection of both the types of barrel architectures in FLD families.

Objective 3 was to understand the evolution of new folds from the ferredoxin-like fold. The inferred evolutionary relationships based on the combined sequence, structure and function study has helped us to understand the origin of novel protein folds from FLDs. Our examples include three major evolutionary mechanisms, *viz.*, domain duplication and fusion, domain swapping, and circular permutation, as responsible for the emergence of novel ferredoxin-like protein structures.

In this study, examples of novel fold emergence by the domain duplication and fusion, and the domain swapping in FLDs included the DN-DC domains of Acr_tran family pumps and LigT-like fold. The study explains the atypical domain swap in TolC docking domain of Resistance-Nodulation-Division pump (RND) families (AcrB, AcrF, AcrD, MexB, MexY, MdtA, MdtB, MdtC, MdtF/YhiV, SwrC, CzcA, CnrA and CusA) resulting in fold change. Typically, the domain swapping results in topologically similar oligomers and retains their monomer fold. Interestingly, the crystal structure of TolC docking domain of a truncated AcrD pump protein (a member of homolog RND pump families) revealed a fully-folded FLD as compared to the swapped domain structure (topologically different structure) in full-length protein structure. The domain-swapped structure of TolC docking domain of RND pump is a highly conserved feature and needed for oligomerization and stability of the tripartite pump assembly. Thus, the study not only provides a rare example of domain swap but also reveals how the core elements of the domain have been involved in structural rearrangement under the functional constraints.

The study also reveals FLD similarity in LigT-like members and suggests the duplication and fusion of swapped FLDs in LigT-like fold. The full-length sequence similarity of swapped FLD1 of LigT along with the conserved histidine residue suggests remote homology relationship with the YbeD-like superfamily. The study also finds a novel example where the ferredoxin-like scaffold has emerged in N0 domain of phage tail and secretins superfamily.

Our remote homology search analysis with FLDs also provided a rare example of a structural switch among homologous proteins. Example includes the ferredoxin-like (alpha+beta domain) and chorismate mutase-like (all-alpha domain) regulatory domains of

DAHP synthase. MSA of both regulatory domain family members was generated and identified conservations of some functionally important residues (inhibitor-binding and catalytic domain interacting residues). The MSA also suggested various hydrophobic side chain residues and polar charged residues conservation through the entire length of the domain. Though, in the absence of global structure similarity it becomes a matter of subjectivity regarding either possible homology (divergence) or possible analogy (convergence) as sequence similarity is only marginal.