Fucose is a major component of complex glycoconjugates and occurs in various organisms ranging from bacteria to mammals (Becker et al. 2003). In humans, L-fucose is a component of various glycoproteins and glycolipids that play roles in recognition, cell-cell interactions and signaling, and thus mediate various biological phenomena such as cell adhesion (T.W. Liu et al. 2009), fertilization and embryonic development (Phopin et al. 2013), pathogenicity and colonization (Hooper et al. 2001). Fucosylated glycoconjugates play a role in various human diseases including cancer metastasis (Li et al. 2006; Shah et al. 2008; Miyoshi et al. 2008; Moriwaki et al. 2010), fucosidosis (Michalski et al. 1999; Sulzenbacher et al. 2004; Osanjo et al. 2007; Alhadeff et al. 1978; Studies et al. 1975) and inflammation (Ali et al. 2008). L-fucose is generally present at the non-reducing end of the glycan. Since, L-fucose is frequently localized at the outer terminal region of the glycan, it is exposed to the surrounding microenvironment and may be recognized by fucose-binding lectins or antibodies.

F-type lectins are fucose-binding lectins with fucose and calcium binding sequence motifs and an F-type lectin fold (Bianchet et al. 2002). Several eukaryotic F-type lectins (e.g. *Anguilla anguilla* agglutinin, *Morone saxatilis* agglutinin) (Bianchet et al. 2002, Bianchet et al. 2010), and the F-type lectin domains (FLDs) of a couple of bacterial proteins (*Streptococcus pneumoniae* SP2159 and *Streptococcus mitis* lectinolysin) have been biochemically and structurally characterized (Boraston et al. 2006; Farrand et al. 2008). Vasta et al. 2004 have previously examined the structural as well as the functional diversity of FLDs in their research articles (Vasta et al. 2004; Odom et al. 2006). However, the extent of occurrence of FLDs and their domain architectures have not been explored much, especially in bacteria. Due to the emergence of Next Generation Sequencing (NGS) as well as whole genome sequencing for

numerous organisms, it is likely that online databases contain many new FLD sequences. Information about proteins containing FLDs in different architectural contexts may give us insights into the biological functions of this domain, the cellular functions FLDs participate in, and the biological activities they direct and perhaps modulate through their FLD. A rigorous survey of FLD containing proteins family would, therefore, be helpful in enhancing our knowledge about the biological function displayed by this domain, especially in proteins containing FLDs in diverse domain organizations. Cytotoxic molecules conjugated to fucose binding antibodies that show high specificity for fucose-containing cell surface antigens is a potential way to specifically target cancer cells. In this regard, the *Streptococcus mitis* FLD that is associated with a cholesterol-dependent cytolysin domain has been recently postulated as an alternate solution (Feil et al. 2012). It is likely that analysis of the various domains co-associated with bacterial FLDs will provide additional strategies of directing biological activities to fucosylated cell surface antigens. Given that L-fucose is a structural constituent of the cell wall and capsule structures of various bacteria, Nod factors (important signaling agents in nitrogen-fixing symbiotic bacteria) and secondary metabolites such as antibiotics, a detailed study of bacterial F-type lectins could be the first step towards such a grand initiative.

Hence, the first objective of this study was to perform bioinformatics analysis to survey the prevalence of the FLD in prokaryotes and eukaryotes. Chapter 2 describes the rigorous survey of FLDs in prokaryotes as well as in eukaryotes by scouring the genomic sequence information in the publically available databases with sensitive bioinformatics tools and sequence search techniques. FLDs were examined for their architectural contexts in different species, for their sequence characteristics, taxonomic

spread, and phylogenetic distribution. Using robust search strategies, 437 FLD sequence clusters were identified from proteins that belonged to eukaryotes, eubacteria, and viruses. FLDs were identified in proteins from numerous forms of life. FLDs were found for the first time in fungi, viruses, birds, reptiles and primitive mammals. FLDs were found to be present in diverse classes of bacteria. As previously suggested by Vasta et al. 2004 FLD distribution was found to be selective as it was absent even among closely related lineages (Vasta et al. 2004). FLDs were more widespread in terms of number in eukaryotic metazoans as compared to prokaryotes which suggests that this domain was horizontally transferred to bacteria after its origin in eukaryotic metazoans. Although eukaryotic and bacterial FLDs clustered separately in the phylogenetic tree, several eukaryotic FLDs were interspersed within the bacterial FLD sequences, which suggests that more than one lateral transfer event might have occurred. New prokaryotic and eukaryotic FLDs were identified in this study having diverse domain organizations. Eukaryotic FLDs occurred generally as multiple repeats and sometimes in co-association with variable domains. In Cephalochordata, in the organism, *Branchiostoma floridae*, a large number of FLDs were identified in diverse domain architectures perhaps relating to the heightened presence of innate immune strategies in this organism. Prokaryotic FLDs generally existed singly or in co-association with diverse domains. The prokaryotic FLDs were found co-associated with domains that included sugar binding domains – Ricin B lectin and various CBMs (CBM6 Agarase, CBM6 Cellulase, CBM6 Xylanase, and NPCBM_assoc) and carbohydrate active enzymatic domains – Glycosyl hydrolase (GH) families, GH16, GH20, and GH98, Alpha-L-fucosidase (GH29), alginate lyase, beta-N-acetylglucosaminidase, glycosyltransferases (GT2/GT92, GT17, and fucosyltransferase) and LicD and other enzymatic domains such as lipase, methyltransferase (families 21

and 23), thiol-activated cytolysin and membrane-bound dehydrogenase. The co-occurring domains varied immensely, suggesting that sugar recognition capability might be harnessed in diverse functional contexts. The different domain architectures of FLD containing proteins suggests that FLDs might have evolved in accordance with specific necessities of organisms present in different ecological niches.

The characterized *Streptococcus mitis* lectinolysin and *Streptococcus pneumoniae* SP2159 possess FLDs that might mediate their colonization and pathogenicity by binding to fucosylated glycoconjugates on host tissues (Boraston et al. 2006; Farrand et al. 2008). Similarly, other FLDs from pathogens or commensals might also bind to fucosylated glycoconjugates on host tissues. In prokaryotes, L-fucose is a constituent of the cell wall, antibiotics, carotenoids and nodulation factors (Takaichi et al. 2001) which act as signaling molecules. It is possible that FLDs might be involved in processes involving these molecules in bacteria. Due to the low content of L-fucose in some bacteria, FLDs may enhance the activity of certain proteins like glycosyltransferases, methyltransferases, and glycosylhydrolases by targeting the enzymatic domain to fucosylated substrates and thereby increasing the effective access of substrates to them. Also, similar to some characterized FLDs like Bryohealin (Yoon et al. 2008) and Ranaspumin-4 (Fleming et al. 2009), a few bacterial FLDs might have evolved towards different monosaccharide specificities according to their ecological niche, specific metabolic requirements and abundance of cellular monosaccharides.

The second objective of this thesis was to explore the effect of the FLD on the enzyme activity of the co-associated domain. In order to explore this aspect, a protein from *Streptosporangium roseum* containing an alpha-L-fucosidase domain (*Sr*Fuc), an Npchm_associated domain (*Sr*Na) and an FLD (*Sr*FLD) was selected for study. The

CBM-like architecture of *S. roseum* containing an FLD in association with a carbohydrate-active enzymatic domain suggested that the FLD might target the activity of the enzymatic domain to fucosylated substrates and additionally or alternatively modulate the enzyme activity itself.

Glycan binding studies for characterization of the FLD of this protein from *S. roseum* was first done (discussed in Chapter 2). *S. roseum* recombinant alpha-L-fucosidase wild type (*Sr*FucNaFLD) was cloned, expressed, purified and used for glycan binding studies. For linkage specificity, recombinant *Sr*Fuc**D244A**NaFLD with its catalytic nucleophile D244 mutated to alanine was expressed, purified and analyzed by glycan micro-array analysis at the Consortium for Functional Glycomics (CFG). *Sr*FucNaFLD was found to bind to histo-blood groups ABH, Lewis antigens and fucosylated glycans in various (1-2, 1-3, 1-4, 1-6) α and β linkages. Glycan micro-array analysis results suggested that *Sr*Fuc**D244A**NaFLD has broad glycan binding specificity. *Sr*FucNaFLD is from the aerobic, Gram-positive, free living, garden soil-borne bacterium *Streptosporangium roseum* DSM4321. Since soil can be a rich medium with decaying plant and animal matter, and is home to a multitude of bacteria, invertebrate animals, and plants, it is possible that the presence of FLDs in a garden soil borne-microorganism such as *S. roseum* is related to the glycan components of plant cell walls. The primary cell wall of higher plants has been reported to have fucosylated oligosaccharides (xyloglucans) and complex branched fucosylated polysaccharides (hemicelluloses) (Carpita et al. 1993). It is thus possible that *Sr*FucNaFLD binds to fucose residues on plant matter, too. ELLA (Enzyme-linked lectin assay) studies done with SrFucD244ANaFLD using biotinylated PAA-α-L-fucose at various pH indicated that the *S. roseum* FLD is active at a wide range of pH (5.6 to 9.2). ELLA studies done with *Sr*Fuc**D244A**NaFLD using biotinylated PAA-α-L-fucose in the presence of EDTA

suggest that FLD binding to biotinylated PAA-α-L-fucose is Ca^{2+} independent, similar to AAA and MsaFBP32. In order to study glycan binding FLD activity in the absence of alpha-L-fucosidase domain and also for determining the role of the NPCBM_ASSOC domain (*Sr*Na), domains of *S. roseum* (*Sr*NaFLD, *Sr*Na and three *Sr*FLD constructs) were cloned, expressed and purified. ELLA studies done with all these protein constructs using biotinylated PAA-α-L-fucose indicate that the lectin activity lies only in the FLD and that the NPCBM_ASSOC (*Sr*Na) has no role in fucose binding. Glycan micro-array analysis also determined that *Sr*NaFLD has similar glycan binding specificity as that of wild-type *Sr*Fuc**D244A**NaFLD. However, unlike wild-type, which was binding equally to all fucosylated glycans belonging to different glycan categories, it was observed that *Sr*NaFLD bound with higher signal intensity to H type-2 glycans as compared to H type-1 glycans and more to difucosylated glycans (Lewis^{b} and Lewis^{y}) than to monofucosylated glycans (Lewis^{a} and Lewis^{x}). This is in accordance with surface plasmon resonance experiments performed in our lab which indicate that the FLD of *S. roseum* does indeed have a higher affinity for H type-2 glycans (unpublished data). For determining the role of the critical residues of the FLD motif (H623, R651, and R658), recombinant mutants (*Sr*FucNaFLD**H623A**, *Sr*FucNaFLD**R651A**, *Sr*FucNaFLD**R658A**) were expressed and purified. Also, a mutant protein where both catalytic nucleophile of the alpha-L-fucosidase domain and critical residues of the FLD motif were mutated to Ala, was expressed and purified. ELLA and glycan array studies of these mutants show loss of lectin activity, indicating that these His and Arg residues of the FLD motif are essential for FLD binding activity.

Having characterized the FLD of the *S. roseum* protein (discussed in Chapter 3), the effect of FLD on the enzymatic activity of the co-associated alpha-L-fucosidase was studied (discussed in Chapter 4). For this, the alpha-l-fucosidase domain was first

characterized and then the effect of the co-associated FLD on the enzymatic activity of alpha-L-fucosidase was observed through biochemical assays done of full length *Sr*FucNaFLD protein (which has FLD and Fuc domains on the same polypeptide, in *cis*), *Sr*Fuc protein, and *Sr*Fuc + *Sr*NaFLD protein (where the FLD and the Fuc domains are on different polypeptides or in *trans*). *S. roseum* alpha-L-fucosidase was first cloned, expressed and purified. Physicochemical properties [optimal temperature, optimal pH, effect of metal ions, kinetic parameters for the synthetic substrate (4-Methylumbelliferyl-α-L-fucopyranoside), IC_{50} values for inhibition of L-fucose and Deoxyfuconojirimycin hydrochloride (DFJHCL)] were studied for both full length alpha-L-fucosidase protein and just alpha-L-fucosidase domain. Both full-length alpha-L-fucosidase protein and just alpha-L-fucosidase domain were found to be optimally active at temperature 37 °C and pH 7.5. Metal ion dependence studies done on just alpha-L-fucopsidase domain of *S. roseum* showed enzymatic activity was enhanced in the presence of 10 mM Ca^{2+} ions and Mn^{2+} ions and unaffected by 10 mM Mg^{2+} and 10 mM EDTA. However, Zn^{2+}, Cu^{2+}, and Ni^{2+} inhibited enzyme activity. There was no effect of higher salt concentration (300 mM NaCl) on alpha-L-fucosidase activity. Since Ca^{2+} has an enhancing effect on *S. roseum* alpha-L-fucosidase enzymatic activity, enzyme kinetics was performed both in the presence and absence of $CaCl_2$ for alpha-L-fucosidase domain. The increased V_{max} for *Sr*Fuc enzymatic activity in the presence of Ca^{2+} confirmed that alpha-L-fucosidase activity was enhanced by Ca^{2+} ions.

Enzyme inhibition studies were done of L-fucose on *S. roseum* alpha-L-fucosidase (*Sr*Fuc) protein. It was observed that L-fucose competitively inhibited the enzymatic activity of *Sr*Fuc and the IC_{50} for L-fucose was 2.62 ± 0.17 µM. Enzyme inhibition studies were done with DFJHCL and *S. roseum* alpha-L-fucosidase (*Sr*Fuc) protein. It was observed that DFJHCL competitively inhibited the enzymatic activity of

*Sr*Fuc and the IC_{50} for DFJHCL was 0.38 ± 0.1316 nM. The higher IC_{50} of L-fucose for the full-length *Sr*FucNaFLD as compared to *Sr*Fuc suggests that L-fucose may be binding to the FLD of *Sr*FucNaFLD due to which higher amount of L-fucose is required for inhibition of the alpha-L-fucosidase domain. However, the IC_{50} values of L-fucose for inhibiting (*Sr*Fuc + *Sr*NaFLD) was also higher in *Sr*Fuc + *Sr*NaFLD as compared to *Sr*Fuc which suggests that higher IC_{50} of L-fucose for inhibiting full-length *Sr*FucNaFLD protein as compared to *Sr*Fuc protein is not a *cis*-acting effect. No significant change in IC_{50} values of DFJHCL for inhibiting full-length *Sr*FucNaFLD, *Sr*Fuc, and (*Sr*Fuc + *Sr*NaFLD) proteins was found, suggesting that FLD may not bind to DFJHCL with comparable or better affinity than the alpha-L-fucosidase domain and thus may not be varying the IC_{50} value.

For substrate preference by alpha-L-fucosidase domain, enzyme kinetics was done in the presence of natural fucosylated glycans and the released L-fucose was detected by FDH-resazurin/diaphorase assay. The effect of the FLD on hydrolysis activity of *Sr*Fuc on natural fucosylated glycans was also studied. The alpha-L-fucosidase domain of *S. roseum* was found to more active against the substrates, Lewis[a] tetraose and Blood group H type-2 with extended motif (Blood group H antigen tetraose type-2, Blood group H antigen pentaose type-2 proparagyl) when co-associated with the FLD in *cis* (on the same polypeptide) but not in *trans* (on different polypeptides). The preferred substrates for the alpha-L-fucosidase domain of *S. roseum* are thus Lewis[a] tetraose (with a Fucα1-4 linkage) and blood group H type-2 with motif pattern [Fucα1-2Galβ1-4GlcNAc]. Fucosylated glycans having (Fucα1-3) linkage could not be identified as a substrate for *Sr*Fuc as they had free L-fucose contamination present due to which enzyme assays performed were not conclusive.

S. roseum is a soil-borne bacterium and so it is likely that besides blood antigens (extended blood group H type-2 with motif pattern and Lewis[a] tetraose) that are preferred as natural substrates for hydrolysis by alpha-L-fucosidase domain, xyloglucan (present in decaying plant matter) may act as a natural substrate. The natural fucosylated glycans used as potential substrates of *S. roseum* alpha-L-fucosidase domain were soluble oligosaccharides. Considering the high rate of diffusion in aqueous media for such substrates, the *cis*-acting enhancer effect of FLD on the enzymatic activity of a co-associated alpha-L-fucosidase domain might be either due to altered (increased) access to substrate facilitated by the relative orientation of the two domains of the protein, or additionally, perhaps through a bind and jump mechanism, whereby the substrate may bind to the FLD first, like in a typical CBM, but then jump to the alpha-L-fucosidase active site. L-fucose released during the reaction may also jump to the FLD binding pocket from the alpha-L-fucosidase domain and the presence of the FLD might thus also serve to lower product inhibition. The glycan binding studies of FLD for *Sr*FucNaFLD, suggest that *Sr*NaFLD binds better to H type-2 glycans as compared to H type-1 glycans. Surface plasmon resonance experiments in our lab also indicate that the FLD of *S. roseum* does indeed have a higher affinity for H type-2 glycans (unpublished data). A similar substrate preference (H type-2 glycans) for the FLD and the alpha-L-fucosidase domain suggests that the FLD might be influencing and/or altering the substrate preference of the *cis*-associated alpha-L-fucosidase domain.

Glycan binding studies were also done on the FLD from *Cyanobium sp.* that was found co-associated with an industrially relevant methyltransferase, FkbM. For this, recombinant *Cyanobium sp.* FkbmFLD was cloned, expressed and purified and used for glycan binding studies at CFG. The glycan binding studies suggest that it has a narrow

and fine specificity for only a subset of complex fucosylated glycans with an extended type-2 blood group H motif, Fucα1-2Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ. This is a very interesting finding since, to our knowledge, no characterized fucose-binding lectin reported in the literature has such a narrow specificity for this category of glycans, and this makes *Cyanobium sp.* FkbMFLD a useful lectin for specific glycan labeling applications.