F-type lectins, also called fucolectins, are typically L-fucose binding lectins with conserved L-fucose binding and Ca^{2+} binding sequence motifs and a conserved F-type lectin structural fold (Bianchet et al., 2002). F-type lectins are a rather recently characterized family of lectins. Identified across almost all genera ranging from microbes to invertebrates to vertebrates, they are present as a single copy or an array of tandemly repeated domains (Ahmed et al., 2008; Vasta et al., 2017). The three-dimensional structure of the *Anguilla Anguilla* agglutinin (AAA) complexed with L-fucose (PDB ID 1K12) revealed a novel fucolectin fold (Bianchet et al., 2002) with the carbohydrate recognition domain being a characteristic β-jellyroll fold possessing the conserved L-fucose binding motif and Ca^{2+} binding sequence motif. The typical L-fucose binding sequence motif as deduced by Bianchet et al (Bianchet et al., 2002) is HX(26)RXDX(4)(R/K) where the conserved His and two Arg residues form H-bonds with the ring oxygen (O5), and hydroxyl groups 4-OH (axial) and 3-OH (equatorial) of the L-fucose molecule.

The eukaryotic F-type lectins had been well studied while little was known about their microbial counterparts before a group in our lab performed extensive bioinformatics analysis using advanced search techniques to mine 437 FLD sequence clusters (clustered at 80% sequence identity) from diverse phylogenetic sources (Bishnoi et al., 2015). This study identified F-type lectin domain-containing proteins from diverse microbial sources. Of the ninety five FLD sequence clusters from prokaryotes identified in the study by Bishnoi et al., almost one third i.e., 34 sequences had typical FLD sequence motifs [HX(26)RXDX(4)R/K], while another one third had conserved residues but variable motif length (Bishnoi et al., 2015). Bishnoi et al. also found 8 FLD sequences with conserved motif length and substitution of key amino acid residues, and 23 FLD sequences with

variation in motif length and substitution of key amino acid residues as well. In these atypical FLDs, polar residues more often replace critical residues, albeit there are also FLDs with apolar residues in these positions. Thus, almost one third of the clustered FLD sequences possess variation in the typical FLD sequence motif with respect to the conserved residues of the FLD motif (Bishnoi et al., 2015). It was demonstrated in *S. mitis* lectinolysin that mutation of the first Arg to Ala (R112A) leads to abrogation of its lectin activity (Farrand et al., 2008a). The other conserved His and Arg, however, have not been similarly experimentally demonstrated to be critical for L-fucose binding. Prior to this thesis, no experimental evidence was presented to suggest if these conserved residue positions (H-R-R) can tolerate the presence of other polar amino acid residues of similar nature and if naturally occurring FLDs with substitutions of one or more of these conserved residues are functional lectins.

Chapter 4, therefore, deals with the biochemical characterisation of 10 FLDs from various microbial sources, including one typical F-type lectin domain with a conserved FLD sequence motif and nine atypical F-type lectin domains possessing variations in the motif length and/or the critical conserved residues. The typical F-type lectin domain included in this study is *Mh*FLD1 from *Myxococcus hansupus* (previously designated as *Myxococcus* sp. contaminant ex DSM 436). The recombinantly produced *Mh*FLD1 was found to bind exclusively to bio-PAA-α-L-fucose and not to other bio-PAA-monosaccharides in enzyme linked lectin assays performed with a panel of biotinylated PAA-linked monosaccharides. The lectin activity of *Mh*FLD1 for PAA-α-L-fucose was inhibited by L-fucose, Lewis^{b}, Lewis^{Y}, type-2 blood group antigen H-pentasaccharide, type-1 blood group antigen H-pentasaccharide, blood group antigen B tetrasaccharide,

and blood group antigen H trisaccharide. These results corroborate with those of the glycan microarray analysis of *Mh*FLD1. The lectin showed selective binding for a number of fucosylated glycans with the motif, Fucα1-2Galβ1-3GlcNAc, and type-1 blood group H motif in glycan microarray analysis.

The FLDs with atypical FLD sequence motifs that we selected for the study were bacterial protein domains described by Bishnoi et al (Bishnoi, R., et al. 2015) - *Arthrospira platensis* FLD (*ApFLD*) with the motif **CX(26)RVDNIDYSER** (with Cys replacing the His residue), *Actinomyces turicensis* FLD (*At*FLD) with the motif **AX(26)RSADDQCNGNQSCAER** (with Ala replacing the His residue), *Campylobacter coli* (*Cc*FLD) with the motif **HX(26)IKDMKFRHR** (with Ile replacing the first Arg residue), *Commensalibacter intestini* (*Ci*FLD) with the motif **LX(26)RLDQNQK** (with Leu replacing the His residue and Lys replacing the second Arg residue), *Methylobacterium extorquens* (*Me*FLD) with the motif **HX(26)SLKDPGMAAR** (with Ser replacing the first Arg residue), *Rhodopirellula baltica* (*Rb*FLD) with the motif **RX(33)RTDPGGPK** (with Arg replacing the His residue and Lys replacing the second Arg residue), *Xenococcus* sp PCC 7305 (*Xp*FLD) with the motif **CX(26)CIDQGEETVR** (with Cys residues replacing the His residue and the first Arg residue), *Candidatus solibacter usitatus ellin* (*Ce*FLD) with the motif **QX(26)RTDCCGSR** (with Gln replacing the critical His residue), and *Saccharomonospora cyanea* (*Sc*FLD) with the motif **TX(26)RVDCCADR** (with Thr replacing the His residue). In *Rb*FLD, *Ce*FLD and *Sc*FLD, the FLD motif length is conserved; in all the other atypical FLDs selected here, there are variations in the length of the FLD motif.

Of the nine atypical FLDs selected, *Actinomyces turicensis* FLD (*At*FLD) with the motif **AX(26)RSADDQCNGNQSCAER** and *Saccharomonospora cyanea* (*Sc*FLD) with the motif **TX(26)RVDCCADR** showed significant binding to PAA-α-L-fucose as determined by

enzyme-linked lectin assay. *Sc*FLD bound to a range of fucosylated glycans including blood group antigen H type 1 and type 2, blood group antigen A, Lewis^x, and Lewis^y antigens on glycan microarray version 5.3. This is the first study to demonstrate glycan binding specificities of FLDs with variations in the typical FLD sequence motif. This study suggests that residues other than His might interact with L-fucose in certain atypical FLDs.

Chapter 6 addresses the query, if FLDs with substitutions of the critical His and/or Arg residues retain lectin activity, and via what alternative residues. To address this question, I adopted two strategies (i) study of a typical L-fucose binding FLD and site-directed mutagenesis of the critical His and/or Arg residue, and (ii) study of naturally occurring atypical FLDs with natural substitutions of critical His and/or Arg residues. For the first strategy, a model F-type lectin (*Sr*NaFLD) from *Streptosporangium roseum*, that has been previously characterized in our lab, as L-fucose binding lectin, was selected. The FLD has the sequence motif, HX(27)RXDX(4)R/K with the conserved L-fucose binding motif residues, His 623, Arg 651 and Arg 658. Site-directed mutagenesis of these conserved residues was performed, mutating them to residues naturally substituting them in atypical FLDs. Besides the wild type *Sr*NaFLD, the H623S, R658K, and D653E mutants of *Sr*NaFLD also bound to biotinylated PAA-α-L-fucose in an L-fucose dependent manner, albeit with weaker binding intensity than wild type *Sr*NaFLD. None of the other mutants bound to biotinylated PAA-α-L-fucose or any other bio-PAA-monosaccharide as assayed by the enzyme linked lectin assay. This suggested that the conserved His of the FLD motif may be replaced by Ser without loss of lectin activity.

In order to execute the second strategy, the atypical FLDs from *Actinomyces turicensis* and *Saccharomonospora cyanea* were selected. *At*FLD has an Ala in the position of the conserved His and an adjacent Ser that seemed likely to be involved in L-fucose binding

instead of the missing His residue. *Sc*FLD has a Thr residue that replaces the conserved His. I used site-directed mutagenesis and ELLA experiments to explore the role of the various L-fucose binding residues in these atypical FLDs and investigate possible compensatory interactions from other residues that enable L-fucose binding in the absence of the defined typical L-fucose binding residues. In *At*FLD, it was found that mutating the Ser to Ala results in significant reduction in the lectin's L-fucose binding ability. The activity, however, was not abrogated completely. I hypothesized that there were certain other interactions involved, which could compensate for the absence of a polar residue like Ser, or His at the active site. Structure modelling using 1K12 as template and analysis of residues lying within 6 Å radius of L-fucose from 1K12 structure identified residues Ser81, Thr46, Glu151, and Asn153 to be present in the 6 Å radius of 1K12 L-fucose. Site directed mutants of the mentioned residues were generated as well. The residue Glu151 lies close to the C-terminus of the FLD, and the position is somewhat conserved in almost all of the FLDs retrieved by Bishnoi et al. Further, Glu151 aligns with the Ser142 of 1K12 which is a part of CDR5. Interestingly, the mutation of Glu151 to either Ser or Ala caused almost complete abolition of L-fucose binding. Therefore, on the basis of experimental data, and structural model studies, I propose that Glu151 participates in L-fucose binding and is perhaps crucial for lectin activity in atypical FLDs, providing compensatory interaction(s) for the loss of conserved His. Site directed mutagenesis of Glu139 in *Sc*FLD again resulted in abrogation of lectin activity, thus reconfirming the role of Glu residue in CDR5 in L-fucose binding. Besides, mutagenesis of Asn141 in *Sc*FLD, which aligns with the His144 of 1K12, also resulted in loss of lectin activity.

The results with *At*FLD and *Sc*FLD Glu151 and Glu139 confirm that this residue in CDR5 definitely plays a critical role in the lectin activity of these FLDs and probably provides

polar end for binding to L-fucose. The Asn153 and Asn141 in *At*FLD and *Sc*FLD align with the His144 of 1K12. His 144 in 1K12 forms a salt bridge with the Arg86 and keeps it in position to interact with L-fucose. From the results with N141A of *Sc*FLD, a similar role is conceived for *At*FLD, too.

To conclude, our study with typical and atypical F-type lectin domains implies that FLDs with Ser or Thr in place of His, Glu in place of Asp, and Lys in place of second Arg of the typical FLD motif are likely to be functional lectins. It can be speculated that FLDs that lack His/Ser/Thr but nevertheless possess a Glu in CDR5 that is equivalent to E151 of *At*FLD are likely functional lectins. Owing to the hypervariability present in the CDR loops, a sequence alignment might not optimally indicate the structurally equivalent residues of the critical residues of the FLD sequence motif. An example is *At*FLD where S42 rather than A43 is structurally equivalent to the critical His residue, H52 of AAA. Therefore, FLDs with an apolar residue such as Ala in place of His, but with a Ser/Thr/His in the vicinity might be functional lectins, too.

To summarize, I find that there is greater plasticity of the L-fucose binding pocket than previously apparent and amino acid residues other than His/Arg/Arg might contact the L-fucose in the active site of the FLD. This study calls for an expansion of the typical FLD motif to include Ser/Thr in place of His and Glu in place of Asp. Based on my current findings and the analysis of FLD sequence motifs in the previous study by Bishnoi et al. (Bishnoi, R., et al. 2015), the FLD sequence motif may be updated to H/S/TX(26-36)RX(0-9)D/EX(1-14)R/K). Further, my study identifies a CDR5 Glu residue that contributes to L-fucose binding in certain atypical FLDs. If future structural experiments confirm that this Glu residue replaces the His/Ser/Thr residue to directly interact with L-fucose, the FLD sequence motif can also be updated to H/S/T/XX(26-36)RX(0-9)D/EX(1-14)R/KX(51-64)E.