ABSTRACT

This thesis consists of two parts. The first part is concerned with further structural and related studies of jacalin, one of the two lectins found in jack fruit seeds. The second part deals with the search of mycobacterial and archeal genomes for lectins.

The β-prism I fold was identified as a lectin fold through the X-ray analysis of jacalin way back in 1996. Subsequent structural studies on jacalin are described in the first chapter in context of the overall efforts on lectins with particular reference to those on lectins with β-prism I fold. The structure of jacalin has been thoroughly characterized through the analysis of several crystals. The extended binding site of the lectin, made up of the primary binding site and secondary sites A and B, has also been characterized through studies on different jacalin-sugar complexes. However, nuances of jacalin-carbohydrate interactions remain underexplored with respect to two specific issues. The first issue is concerned with the structural basis for the lower affinity of jacalin for β-substituted sugars. The second has to do with the influence of the anomeric nature of the glycosidic linkage on the location of the reducing and non-reducing sugars in disaccharides when interacting with jacalin. Part of the work described in the thesis addresses these two issues.

It was surmised that the lower affinity of β-galactosides to jacalin as compared to α-galactosides, is caused by steric interactions of the substituents in the former with the protein. This issue is explored both energetically and structurally in Chapter 2 using appropriately derivatized monosaccharide complexes of jacalin. It turns out that the earlier surmise is not correct. The interactions of the substituent with the binding site remain essentially the same irrespective of the anomeric nature of the substitution. This is achieved through a distortion of the sugar ring in β-galactosides. The difference in energy, and therefore affinity, is caused by the distortion of the sugar ring in β-galactosides. The elucidation of this unprecedented distortion of the ligand as a strategy for modulating affinity is of general interest. The crystal structures also provide a rationale for the relative affinities of the different carbohydrate ligands to jacalin.

The crystal structures of jacalin complexed with α-linked oligosaccharides Gal α-(1,4) Gal and Gal α-(1,3) Gal β-(1,4) Gal, as described in Chapter 3, have been determined with the primary objective of exploring the effect of linkage on the location of reducing and non-reducing sugars in the extended binding site of the lectin, an issue which has not been studied thoroughly. Contrary to the earlier surmise based on simple steric considerations, the two structures demonstrate that α-linked sugars can bind to jacalin with non-reducing sugar at the

primary binding site. This is made possible substantially on account of the hitherto underestimated plasticity of a non-polar region of the extended binding site. Modelling studies involving conformational search and energy minimization, along with available crystallographic and thermodynamic data, indicate a strong preference for complexation with Gal β-(1,3) Gal with the reducing Gal at the primary site, followed by that with Gal α-(1,3) Gal, with the reducing or non-reducing Gal located at the primary binding site. This observation is in consonance with the facility of jacalin to bind mucin type O-glycans containing T-antigen core.

Crystal structures of jacalin in complex with GlcNAc β-(1,3) Gal-β-OMe and Gal β-(1,3) Gal-β-OMe have also been described in Chapter 4. The binding of the ligands to jacalin is similar to that of analogous α-substituted disaccharides. However, the β-substituted β-(1,3) linked disaccharides get distorted at the anomeric centre and the glycosidic linkage. The distortion results in higher internal energies of the ligands leading to lower affinity to the lectin. This confirms the possibility of using ligand distortion as a strategy for modulating binding affinity. Unlike in the case of β-substituted monosaccharides bound to jacalin, where a larger distortion at the anomeric centre was observed, smaller distortions are distributed among two centres in the structures of the two β-substituted β-(1,3) linked disaccharides presented here. These disaccharides, like the unsubstituted and α-substituted counterparts, bind jacalin with the reducing Gal at the primary binding site, indicating that the lower binding affinity of β-substituted disaccharides is not enough to overcome the intrinsic propensity of Gal β-(1,3) Gal based disaccharides to bind jacalin with the reducing sugar at the primary site.

Although originally isolated from plants, lectins were also found subsequently in all forms of life, including bacteria. Studies on microbial lectins have not been as extensive as on those from plants and animals, although there have been some outstanding individual investigations on bacterial toxins like ADP-ribosylating toxins and neurotoxins. In addition to bacterial toxins, adhesins, β-trefoil lectins and cyanobacterial lectins form other important subgroups which have been explored using crystallography. Features pertaining to their three dimensional folds, carbohydrate specificity and biological properties are described in Chapter 5, to set the stage for the work discussed in the second part of the thesis. Studies on mycobacterial lectins were unexplored until work was initiated in the area in this laboratory some years ago. One of the lectins, identified on the basis of a bioinformatics search of *M. tuberculosis* H37Rv genome was cloned, expressed and crystallized. Also cloned, expressed and crystallized is another lectin from *M. smegmatis*. Biophysical and modelling studies were

carried out on the full length protein containing this lectin. However, systematic efforts on mycobacterial lectins were conspicuous by their absence. The first chapter (Chapter 6) in the second part of the thesis is concerned with a genomic search for lectins in mycobacterial genomes. It was also realized that hardly anything is known about archeal lectins. Therefore, as discussed in the final chapter, a genomic search for archeal lectins was undertaken.

Sixty-four sequences containing lectin domains with homologs of known three-dimensional structure were identified through a search of mycobacterial genomes and are described in detail in Chapter 6. They appear to belong to the β-prism II, the C-type, the *Microcystis virdis* (MV), and the β-trefoil lectin folds. The first three always occur in conjunction with the LysM, the PI-PLC, and the β-grasp domains, respectively while mycobacterial β-trefoil lectins are unaccompanied by any other domain. Thirty heparin binding hemagglutinins (HBHA), already annotated, have also been included in the study although they have no homologs of known three-dimensional structure. The biological role of HBHA has been well characterized. A comparison between the sequences of the lectin from pathogenic and nonpathogenic mycobacteria provides insights into the carbohydrate binding region of the molecule, but the structure of the molecule is yet to be determined. A reasonable picture of the structural features of other mycobacterial proteins containing one of the four lectin domains can be gleaned through the examination of homologous proteins, although the structure of none of them is available. Their biological role is yet to be elucidated. The work presented here is among the first steps towards exploring the almost unexplored area of the structural biology of mycobacterial lectins.

As mentioned in Chapter 7, forty six lectin domains, which have homologues among well established eukaryotic and bacterial lectins of known three dimensional structure, have been identified through a search of 165 archeal genomes using a multi-pronged approach involving domain recognition, sequence search and analysis of binding sites. Twenty one of them have the 7-bladed β-propeller lectin fold while 16 have the β-trefoil fold and 7 the legume lectin fold. The remainder assumes the C-type lectin, the β-prism I and the tachylectin folds. Acceptable models for almost all of them could be generated using the appropriate lectins of known three dimensional structure as templates, with binding sites at one or more expected locations. The work represents the first comprehensive bioinformatic study of archeal lectins. The presence of lectins with the same fold in all domains of life indicates their ancient origin well before the divergence of the three branches. Further work is necessary to identify archeal lectins which have no homologues among eukaryotic and bacterial species.

The work presented in the thesis has been reported in the following publications.

1. Abhinav, K.V., Sharma, A., Vijayan, M. (2013). Identification of mycobacterial lectins from genomic data. *Proteins,* **81**, 644-657.

2. Abhinav, K.V., Sharma, K., Swaminathan, C.P., Surolia, A., Vijayan, M. (2015). Jacalin-carbohydrate interactions. Distortion of the ligand molecule as a determinant of affinity. *Acta Cryst*., **D71**, 324–331.

3. Abhinav, K.V., Samuel, E., Vijayan, M. (2016). Archeal lectins. Identification through a genomic search. *Proteins,* **84**, 21-30.

4. Abhinav, K.V., Sharma, K., Surolia, A., Vijayan, M. (2016). Effect of linkage on the location of reducing and non-reducing sugars bound to jacalin. *IUBMB Life*, **68**, 971-979.

5. Abhinav, K.V., Sharma, K., Surolia, A., Vijayan, M. (2016). Distortion of the ligand molecule as a strategy for modulating binding affinity. Further studies involving complexes of jacalin with β-substituted disaccharides. *IUBMB Life* (In press).

In addition, the candidate is a co-author in the following publications:

6. Chetnani, B., Kumar, P., Abhinav, K.V., Chibber, M., Surolia, A., Vijayan, M. (2011). Location and conformation of pantothenate and its derivatives in *Mycobacterium tuberculosis* pantothenate kinase: insights into enzyme action. *Acta Cryst*., **D67**, 774–783.

7. Abhinav, K.V. and Vijayan, M. (2014). Structural diversity and ligand specificity of lectins. The Bangalore effort. *Pure and Applied Chemistry,* **86**, 1335-1355.