CONCLUSIONS

In this study an attempt has been made to analyse various biomolecular databases in order to throw more light on structural and conformational aspects of carbohydrates, free as well as linked to proteins.

An analysis of the Complex Carbohydrate Structure Database (CCSD), the only available database for oligosaccharide structures, has been carried out to find out the distribution and linkage pattern of the most commonly occurring sugars, particularly six carbon sugars. This analysis attempts for the first time to quantify the information contained in these structures and reveals many interesting features. The aldohexose sugars, glucose, galactose and mannose, occur in large numbers in this database. For glucose and galactose the occurrence of β-anomer is more than that of the α-anomer but for mannose the occurrence of α-anomer is higher than that of β. The ratio of α, β anomers of glucose, galactose and mannose in the oligosaccharide structures corresponds significantly to the α, β equilibrium values for these sugars in solution indicating that the enzymes involved in the biosynthesis of these structures might use an unaltered solution pool of sugar residues. For fucose and rhamnose the α-anomeric configuration predominates. The sugars, α-D-Neup5Ac, α-L-Fucp, α-L-Araf, α-D-Neup5Gc, α-D-Xy1p and α-D-Galp, occur preferentially at the non-reducing terminal end of the oligosaccharide
structures and seem to have been selected to play a role in interaction with other molecules. There are a large number of structures containing Galβ(1-4)GlcNAc disaccharide linkage. The axial-equatorial linkage is dominant in homo disaccharides while the equatorial-equatorial linkage is common in hetero disaccharides.

The frequency of occurrence of the amino acid X in the consensus glycosylating sequence, Asn-X-Ser/Thr, in proteins and the conformational and hydrogen bonding features have been analysed using the PDB database. 488 non homologous proteins bearing 696 Asn-X-Ser/Thr (X not Pro) sequences were analysed. More than 65% of Asn, when they occur as part of the consensus sequence lie on the surface of the protein, implying a potentiality for glycosylation. At position X in the consensus sequence segment, the amino acids Gly, Asn and Phe are preferred as indicated by the statistically significant positive DP (Deviation Parameter) values. The high value of DP for Asn is due to the preferential occurrence of homodoublets while for Phe it is due to stacking interaction of the aromatic ring with the glycan. Gly at X position in the consensus glycosylating sequence is functionally significant due to its preference and its high percentage of occurrence in proteins. The Ramachandran (Φ, Ψ) angles around Gly in the consensus sequence show clustering in the region which is disallowed for non-glycyl residues. In this region a hydrogen bond between the side chain of Asn
and the peptide backbone/side chain of Ser/Thr is possible reflecting the positional as well as a conformational role of Gly containing glycosylating sequences. In most of the cases a direct or water mediated hydrogen bond between OD1 of Asn and OG of Ser/Thr is possible reflecting the possible importance of these hydrogen bonding in the N-glycosylation process. This paves a way to identify Gly containing glycosylatable sequences in proteins.

Statistical analysis carried out on the O-glycosylated sequences from the O-GLYCBASE database indicates that there is a pronounced positional preference for certain amino acids at various positions around the O-glycosylation site. Pro occurs preferentially at many positions close to the site of glycosylation and, in particular, it is strongly favoured at -1 and/or +3 positions relative to the O-glycosylated Ser/Thr, irrespective of single and multiple glycosylation sites. This indicates that proline plays a structural role in directing O-glycosylation in contrast to the negative role it plays in N-glycosylation. Proline more frequently occurs at the -2 and +2 positions when the site of glycosylation is a Ser residue. Around the multiple glycosylation sites Ser and Thr are preferred, probably due to the effect of clusters of closely spaced O-glycosylation sites. The other amino acids preferred favourably around multiple glycosylation sites are Ala, Gly, Asp, His and Val. Cysteine, aromatic amino acids and amino acids with bulky side chains hinder O-glycosylation. In mucin type O-glycosylation the acidic amino
Acids Asp and Glu are also preferred at various positions around Ser/Thr indicating the possible structural role of these amino acids in the mucin type glycosylation. The preference of certain potential sequence motifs of O-glycosylation have been identified. The conformation of the sugar linked to the hydroxyamino acids indicates the formation of direct or water mediated hydrogen bonding between the glycan and the peptide backbone/side chain, and this may act as one of the stabilising forces.

The carbohydrate-protein interactions in glycoproteins and glycan-binding proteins have been studied using the Brookhaven Protein Data Bank, particularly the hydrophobic and stacking interactions between the sugars and the amino acids. The hydrophobic interaction between the hydrophobic plane in sugars, formed by three axial hydrogens in the stable chair conformation (pointing the same direction in β- sugars of Glc, Gal, Man), or the methyl group of acetamido sugars (NeuNac, GlcNAc, GalNAc) and the hydrophobic/aromatic amino acids were computed. It is noted that in more than 50% proteins there is stacking/hydrophobic interaction between the glycan and the hydrophobic/aromatic amino acids. The angle between the stacking planes is in the range of 30°±10° or 80°±10° indicating the specificity of stacking interaction. The methyl group in acetamido sugars is also taking part in hydrophobic interaction with the hydrophobic/aromatic amino acids. To find the potentiality of various amino acids involved in hydrophobic interaction a
quantity called hydrophobic interaction index (HII) is computed for the two
types of hydrophobic interaction. HII is maximum for Trp in both glycoproteins
and glycan-binding proteins when the hydrophobic sugar plane is involved in
the interaction. When the methyl group of acetamido sugars is involved HII is
maximum for Trp and Leu in glycoproteins and glycan-binding proteins
respectively.

This is a potentially important study and such analyses will surely
contribute an important part of our knowledge base in the future on the
structure and conformation of the carbohydrates as well as proteins to which
they are attached in the process of glycosylation. These results will be of
interest to biophysicists, molecular biologists, bioinformatists, biotechnologists
and protein engineers.