Chapter 1

Introduction

1.1 Carbohydrates

Carbohydrates or saccharides are essential components of all living organisms and are the most abundant class of biomolecules found in nature. The word *saccharide* is derived from the greek word *sakcharon* which means sugar. The name carbohydrate itself distinctly indicates that they are hydrates of carbon (Paulson, 1989; Varki, 1993; Flitsch and Ulijn, 2003). Most often they have the molecular formula $C_n(H_2O)_m$. Green plants, algae, and some bacteria capture the light energy from the sun and store it in the form of carbohydrates during photosynthesis. Breakdown of carbohydrates provides the energy that sustains animal life. In turn, carbohydrates are the metabolic precursors of virtually all other biomolecules. They are covalently linked with a variety of other biomolecules like lipids and proteins and called glycolipids and glycoproteins respectively. Glycolipids and glycoproteins are the major components of the outer surface of mammalian cells and they together are called glycoconjugates. They are also common components of biological cell membranes where they play a variety of biological roles including fertilization, immune defense, viral replication, parasitic infection, structural integrity, cell growth, cell–cell communication, cell–cell adhesion, cell–cell recognition, degradation of blood clots, and inflammation.
(Karlsson, 1995; Crocker and Feizi, 1996). Cell surface carbohydrates can act as receptors for many viruses, bacteria, lectins, and toxins (Varghese et al., 1992; Kelm and Schauer, 1997; Varki, 1997) and they are recognized by the pathogens like bacteria and viruses toward entry into the cell to inflict infections. Hence the cell surface carbohydrates are called as **Molecules of Molecular Recognition.** Carbohydrates are generally classified into monosaccharides, disaccharides, oligosaccharides and polysaccharides.

### 1.2 Monosaccharides

**Monosaccharides** are the fundamental units of carbohydrates since they cannot be divided in to further units. These simple sugars are classified on the basis of the nature of their carbonyl (COO⁻) group and the number of carbon atoms present. If the carbonyl group is an aldehyde (−CHO), the sugar is named as aldose and if ketone (−C=O), the sugar is named as ketose. The smallest monosaccharide has three carbon atoms and named as triose. If the number of carbon atoms are four, five, six, seven, etc., then the sugars are named as tetrose, pentoses, hexoses, heptoses, etc., respectively. Glucose is the most common monosaccharide which is taken as example to illustrate many of the structural features of all monosaccharides. Some of the other biological significant monosaccharides are D–glyceraldehyde, D–ribose, D–mannose, D–galactose and α–D–N–Acetyl neuraminic acid (sialic acid).

### 1.3 Stereochemistry of Carbohydrates

**Stereochemistry** is a branch of chemistry which explains how the atoms of molecule are arranged in a three dimensional space. The three dimensional
arrangements of atoms of a molecule can be easily understood from physical molecular models (balls and sticks) and computer molecular models. But the printed and hand–drawn representations of molecules are two dimensional images that are meaningful only to those who understand with the knowledge of visualizing the third dimension (Hoffmann and Laszlo, 1991).

Carbohydrates are most often represented by two different ways of projections, namely, Fischer projection and Haworth projection. In 1891, Hermann Emil Fischer formulated a two–dimensional representation of a three dimensional organic molecule by projection; hence it is called *Fischer projection*. Initially, Fischer projection was developed to display the stereochemical relationships among carbohydrates with several chiral centers (Lichtenthaler, 1992; Maehr, 1992). In this projection diagram, molecules are represented by crossing vertical and horizontal lines, with each intersection representing a carbon atom. In a Fischer projection, all horizontal bonds project toward the viewer, while vertical bonds project away from the viewer (Carroll and Noble, 1998).

In a molecule, if a carbon is surrounded by four different chemical groups in a three dimensional space, that carbon is called asymmetric or *chiral* carbon. *Enantiomers* are chiral molecules that are mirror images of each other and they are non–superimposable on one another. Generally, minor differences in physical and chemical properties of pure enantiomer compounds which have a single chiral carbon atom can be observed. However, there is one physical property called *optical activity* which is markedly different. Optical activity is a macroscopic physical phenomenon that refers to the ability of a solution of an enantiomer to rotate the plane of polarization of plane–polarized light. One of a pair of
enantiomers rotates the plane of polarization in a clockwise direction (right) hence it is called as **dextrorotatory** (D) and it is denoted by the symbol plus (+). The other rotates the light counterclockwise (left) hence it is called as **levorotatory** (L) and it is denoted by the symbol minus (−). Enantiomers rotate plane polarized light with the same magnitude of degrees, but in opposite directions. In Monosaccharides, if the hydroxyl group on the highest numbered asymmetric carbon (the asymmetric carbon farthest from the carbonyl carbon) is on the right in a Fischer projection, then that monosaccharide is designated as D and if it is on left, that monosaccharide is designated as L. However the designation D or L is merely used to identify the configuration of a given molecule and doesn’t used to specify the sign of rotation of plane–polarized light. If the sign of optical rotation is to be specified in the name, the convention of D or L designation may be used along with a plus (+) or minus (−) sign. Thus D-glucose can also be named as D(+)-glucose since it is dextrorotatory, while D-fructose is levorotatory, can also be named as D(−)-fructose. D-forms of sugars are predominant in nature (Garrett and Grisham, 2010; Cox, 2013) and Fischer projection diagrams for D-forms of Glucose, Fructose and Galactose are shown in figure 1.1(a), (b) and (c) respectively.
Molecules having same molecular formulae but different structures are called **isomers**. Generally if a molecule has \( n \) chiral centers, then it can have \( 2^n \) stereoisomers. In aqueous solution, all monosaccharides with five or more carbon atoms in the backbone occur predominantly as cyclic or ring structures in which the carbonyl group has formed a covalent bond with the oxygen of a hydroxyl group along the chain. The formation of these ring structures is the result of a general reaction between alcohols and aldehydes or ketones to form derivatives called **hemiacetals** or **hemiketals** (figure 1.2(a) and (b) respectively), which contain an additional asymmetric carbon atom and thus can exist in two stereoisomeric forms. If a sugar has a six-membered ring, then it is called **pyranose** in analogy with **pyran** (figure 1.3(a)), similarly, sugars having five-membered rings are called **furanoses** in analogy with **furan** (figure 1.3(b)). The cyclic forms of glucose and fructose with six- and five- membered rings are therefore called as **glucopyranose** and
Figure 1.2 Chemical reactions of alcohols with (a) aldehydes to form hemiacetals and (b) ketones to form hemiketals.

Figure 1.3 Ring structures of sugars, (a) Pyran (six-member carbon ring) and (b) Furan (five-member carbon ring).
fructofuranose respectively (Voet and Voet, 1995; Garrett and Grisham, 2010; Cox, 2013).

Configuration and conformation are the two different and important terms used in understanding the spatial arrangement of atoms or groups of atoms in biomolecules. Configuration of a molecule defines the position of chemical groups around one or more non-rotating bonds or around chiral centers. A configurational isomer cannot be turned into another configuration without breaking covalent bonds. Isomers of monosaccharides that differ only in their configuration about that carbon atom are called anomers, designated as α or β and the carbonyl carbon is thus called anomic carbon. For example, D-glucose can have two different configurations which are α-D-glucopyranose and β-D-glucopyranose as shown in figure 1.4. In D-Glucose, when the hydroxyl group at the anomic carbon is on the same side of Fischer projection as the oxygen atom at the highest numbered asymmetric carbon, the configuration at the anomic carbon is α, as in α-D-glucopyranose. When the anomic hydroxyl is on the opposite side of the Fischer projection, the configuration is β, as in β-D-glucopyranose. However, Fischer projection technique lags when representing the cyclic structures of sugars. Hence a new two dimensional representation of a three dimensional cyclic structure was introduced by Haworth and this projection method is called Haworth projection. Haworth projection is widely used conveniently for representing cyclic structures of monosaccharides (Barker and Baggett, 1978). In cyclic D-sugars, if anomic hydroxyl group occurs below the ring, then it is called α-anomer and if the anomic hydroxyl group occurs above the ring, it is called β-anomer (figure 1.5). Generally different configurations of a molecule have different optical activities and
form different crystalline structures. Fisher and Haworth projection diagrams for 
\(\alpha\)-D-glucopyranose and \(\beta\)-D-glucopyranose are shown in figure 1.4 and 1.5 respectively (Carroll and Noble, 1998; Garrett and Grisham, 2010; Cox, 2013).

Figure 1.4 Fischer projection diagram of D-glucose, (a) \(\alpha\)-D-glucopyranose as \(\alpha\)-anomer and (b) \(\beta\)-D-glucopyranose as \(\beta\)-anomer
**Figure 1.5** Haworth projection diagram of D–glucose, (a) α-D-glucopyranose as α-anomer and (b) β-D-glucopyranose as β-anomer

The word *conformation* was first coined by the British carbohydrate chemist Sir Norman Haworth to describe the various shapes of molecules (Haworth, 1929). This word is now commonly used to describe the different spatial arrangements of atoms or atomic groups produced by rotations about the bonds, without breaking the bonds. The term conformation generally refers to one of several different spatial arrangements that a molecule can achieve by rotations about single bonds between atoms (Elie, 1975; Schneider and Werner, 1992). If a molecule has several bonds about which rotation can occur, then the shape of the molecule can change significantly (Schneider and Werner, 1992). Even though they cannot be isolated, conformations that correspond to energy minima are known as *conformers*, a contraction of conformational isomers. Different conformations for a molecule can be generated without breaking the bonds by rotating the chemical groups about the bonds in order to have reasonable molecular structures without stereochemical clashes. Of the many possible conformations, there is usually a *favoured conformation* that is the conformer with the global minimum energy. The
various factors that contribute to the favoured conformation involved bond-angle strain, bond-torsional strain, steric interactions of the atoms and groups, dipolar interactions, hydrogen bonding effects, solvation effects and hydrophobic interactions (Robyt, 2012).

A chair and a boat are the two possible strain-free conformations of a six-membered carbon ring. They are interconvertible by various rotations of the C—C bonds and they are referred to as conformers rather than isomers. The presence of an oxygen atom in the ring structures denotes that there are two possible chair structures designated by Reeves as 1C1 and 1C (Reeves, 1954), which are interconvertible by rotation of the bonds, passing through a boat structure (figure 1.6). The C1 chair is also called 4C1 chair and 1C chair is called 1C4 chair. In the chair form, the substituents on the ring carbon atoms have two orientations which are axial and equatorial. Axial bonds are nearly perpendicular to the average plane of the ring, whereas equatorial bonds are nearly parallel to this plane. In other words, axial bonds are parallel to the axis of the ring, while equatorial bonds are perpendicular to the axis of the ring and lie along the equator of the chair (figure 1.7). Ring axis is an axis passing perpendicular to the plane of the ring. One chair form can be converted to another by rotating around C—C bonds that changes the equatorial substituents to axial position or vice versa (figure 1.8). The chair conformation is more stable than the boat conformation because of the steric overlap at some positions introduced by the boat conformation (Berg et al., 2006; Philip, 2009; Garrett and Grisham, 2010; Cox, 2013; Rao, 2013).
Figure 1.6 Chair and boat conformations of sugars, (a) C1($^1$C$_1$)–chair form, (b) Boat form and (c) 1C($^1$C$_4$)–chair form

Figure 1.7 Axial and equatorial representations of various conformations of sugars, (a) C1($^1$C$_1$)–chair conformation (b) 1C($^1$C$_4$)–chair conformation and (c) Boat conformation
**Figure 1.8** Two different chair conformations of β-D-glucose, (a) C1(1C1)–chair conformation and (b) 1C(1C4)–chair conformation

### 1.4 Disaccharides

Disaccharides are formed when two monosaccharides are bound together by an O–glycosidic linkage which is formed when a hydroxyl group of one sugar reacts with the anomic carbon of the other. In describing disaccharides or polysaccharides, the end of a molecular chain with a free anomic carbon (the one not involved in a glycosidic bond) is generally called the **reducing end** and such a disaccharide is called as **reducing disaccharide**.

Celllobiose, gentiobiose, maltose, isomaltose and lactose are a few examples of reducing disaccharides. The disaccharide maltose has two D–glucose residues linked by a glycosidic linkage between C1 (anomeric carbon) of one glucose residue and C4 of the other (figure 1.9). Since the disaccharide retains a free anomeric carbon (C1 of the glucose residue on the right), maltose is a reducing disaccharide. Whereas, in a disaccharide, anomeric carbons of both the sugar units are involved in a glycosidic bond, then there is no free anomeric carbon in the disaccharide that is called as **non reducing disaccharide** since it cannot take the linear form. Sucrose is a disaccharide which consists of glucose and fructose and it is formed by plants
but not by animals. In contrast to maltose, sucrose has no free anomeric carbon atom since the anomeric carbons of both monosaccharide units are involved in the glycosidic bond. The substituted anomeric carbons cannot be converted to the aldehyde configuration and thus cannot participate in the oxidation–reduction reaction which is a characteristic feature of reducing sugars. Hence sucrose is a non-reducing disaccharide and it is shown in figure 1.10. Sucrose and trehalose are examples of non-reducing disaccharides. Non-reducing disaccharides are also named as **glycosides** (Garrett and Grisham, 2010; Cox, 2013).
Figure 1.9 Maltose as a reducing disaccharide

Figure 1.10 Sucrose as a non reducing disaccharide
1.5 Oligosaccharides

Oligosaccharides are formed when three to ten monosaccharides are bound together by glycosidic linkages. Oligosaccharides can have linear, cyclic or branched chain structures. Generally they can covalently link with proteins and lipids to form glycoproteins and glycolipids respectively. Due to the inherent flexibility at the glycosidic linkages, oligosaccharides can exist in multiple conformational states in biological environment. Cycloamyloses are interesting and useful group of oligosaccharides which are having cyclic structures. They can form molecular “pockets” of various diameters in solution. These pockets are surrounded by the chiral carbons of the saccharides themselves and are able to form stereo specific inclusion complexes with chiral molecules that can fit into the pockets. Thus, mixtures of stereoisomers of small organic molecules can be separated into pure isomers on columns of cycloheptaamylose, for example. Stachyose is typical of the oligosaccharide components found in substantial quantities in beans, peas, bran, and whole grains. These oligosaccharides are not digested by stomach enzymes, but are metabolized readily by bacteria in the intestines. Oligosaccharides also occur widely as components of antibiotics derived from various sources. Carbohydrate-containing antibiotics show antitumor activity and one of the best examples is bleomycin A2, which is used clinically against certain tumors.

1.6 Polysaccharides

Polysaccharides are formed when more than ten monosaccharides are bound together by glycosidic linkages. Polysaccharides consist of long chain of monosaccharides, usually more than ten monosaccharide units and/or their derivatives. The most common constituent of polysaccharides is D-glucose, but D-
fructose, D-galactose, L-galactose, D-mannose, L-arabinose, and D-xylose are also common. Polysaccharides differ not only in the nature of their component monosaccharides but also in the length of their chains and in the amount of chain branching that occurs. Although a given sugar residue has only one anomeric carbon and thus can form only one glycosidic linkage with hydroxyl groups of other molecules, each sugar residue carries several hydroxyls, one or more of which may be an acceptor of glycosyl substituents. This ability to form branched structures distinguishes polysaccharides from proteins and nucleic acids, which occur only as linear polymers.

1.7 Conformational parameters of carbohydrates

Conformational analysis of carbohydrates is the most important step in understanding the structure and functions of carbohydrates. The overall conformation of an oligosaccharide is greatly influenced by the conformational preferences around glycosidic linkages between successive monosaccharides. The conformational flexibility of carbohydrates can be explained well by taking a disaccharide as example. The rotational degrees of freedom around a glycosidic linkage is higher than other functional groups in a disaccharide system. A simple disaccharide unit (celllobiose) is shown in figure 1.11 with its glycosidic torsional angles φ (phi) and Ψ (psi). The glycosidic torsional angles φ and Ψ are calculated among the four sets of atoms H1–C1–O4–C4 and C1–O4–C4–H4 respectively.
Figure 1.11 Cellulobiose as a typical disaccharide with its glycosidic torsions $\phi$ and $\Psi$
1.8 Sialic acid – A Significant Carbohydrate

During 1930s and 1940s, several acidic sugars were discovered independently by the scientists, Ernst Klenk, Gunner Blix and Gottschalk (Blix et al., 1957). They named these unique sugars differently. However, in early 1950s, the structural elucidation studies revealed that the discovered novel carbohydrates have structural similarity. In order to avoid the confusion that may have arisen in the field of carbohydrate chemistry due to the discovered sugars, three above researchers together called the unsubstituted acidic amino sugar as neuraminic acid (Neu). In 1935, the German scientist, Ernst Klenk isolated a sugar from brain glycolipids (neuro+amine+acid) and hence he called it “neuraminic acid”. In the next year 1936, the Swedish biochemist Gunnar Blix isolated a sugar from salivary gland mucin (Greek meaning of saliva is sialos) since this sugar was christened sialic acid (Sia) by him and that was universally accepted. Neuraminic acid has more than 50 diversified forms and these derivatives of neuraminic acid are collectively called sialic acids (Sias). The structure, chemistry and biosynthesis of sialic acids were elucidated by Gunnar Blix and Ernst Klenk in 1957 (Blix et al., 1957).

Sialic acids are a family of nine carbon acidic monosaccharides usually found at the end of oligosaccharide chains that are attached to glycoproteins and glycolipids (Varki, 1992a; Schauer et al., 1995; Schauer and Kamerling, 1997). They are commonly occupying terminal, non-reducing positions of oligosaccharide chains connected to the cell surface (Varki and Varki, 2007). Sialic acids potentially inhibit many intermolecular and intercellular interactions by virtue of their
negative charge and terminal location (Varki, 1997). In contrast to their role as inhibitors, sialic acids can also be important components of ligands for various recognition phenomena involving carbohydrate binding proteins (Cohen et al., 1983; Wiley and Skehel, 1987; Mandal, 1990; Varki, 1992b). Usually, unsubstituted form of neuraminic acid does not exist in nature. Structural diversity of sialic acid arises from various substitutions at the 4th, 5th, 7th, 8th and 9th carbon positions (Schauer, 1991; Varki, 1992a; Kitazume et al., 1996; Reuter and Gabius, 1996). Generally the amino group present in the 5th position of neuraminic acid is acetylated which forms N-acetyl-neuraminic acid (Neu5Ac) and Neu5Ac is believed to be the biosynthetic precursor for all other members of sialic acid family (Rosenberg and Schengrund, 1976; Schauer, 1982). Hydroxylation of this N-acetyl group gives N-glycolyl-neuraminic acid (Neu5Gc) (Jourdian and Roseman, 1962; Schoop et al., 1969; Roseman, 1970). The amino group (5th position) of neuraminic acid can also be replaced by a hydroxyl group (OH) which leads to 2-Keto-3-Deoxy-Nonulosonic acid (KDN) or de-aminated–neuraminic acid (Nadano et al., 1986; Kanamori et al., 1990). Neu5Ac, Neu5Gc and KDN are shown in figure 1.12 (a), (b) and (c) respectively. Other sialic acids can be acquired from substitution of one or more of the hydroxyl groups of Neu5Ac, Neu5Gc or KDN with acetyl, methyl, lactyl, phosphate or sulphate groups (Warren, 1964; Schauer, 1981; Schauer, 1987; Manzi et al., 1990; Traving and Schauer, 1998). In the still growing family of sialic acids with more than 50 different derivatives of neuraminic acid, Neu5Ac, Neu5Gc and KDN are considered to be significant due to their major roles in various biological processes. Among the three sugars, Neu5Ac is the most significant sialic acid and commonly found in mammalian cells.
Figure 1.12 Diversity of Sialic acids, (a) N-acetyl-neuraminic acid (Neu5Ac), (b) N-glycolyl-neuraminic acid (Neu5Gc) and (c) 2-Keto-3-Deoxy-Nonulosonic acid (KDN)

Neu5Ac is also found in the terminal position of glycoproteins present in the cell surface and acts as a receptor for influenza viruses by allowing initial attachment to viral hemagglutinin. The remarkable feature of Neu5Ac is its negative charge given by carboxylate group (COO⁻) present at C1 position. In aqueous state, Neu5Ac always prefers 2C₅ chair conformation predominantly (Flippen, 1973; Veluraja and Rao, 1980; Rao et al., 1998; Spíwok and Tvaroška, 2009). We have limited evidence for the presence of Neu5Gc in healthy human cells whereas it is found in human cancer cells (Inoue et al., 2010; Malykh et al., 2001; Varki, 2001). But Neu5Gc is commonly occurring in various animal species especially in porcine tissues (Malykh et al., 1998; Morimoto et al., 2001). KDN is found in rainbow trout fish (eggs), teleost fish (sperms and eggs) and also found in lower and higher vertebrates including mammals (Knirel et al., 1989; Strecker et al., 1992; Varki, 1992a; Ziak et al., 1996).
Sialic acids generally exist in four major configurations which are α(2→3), α(2→6), α(2→8) and α(2→9). Among the four, Neu5Ac forms α(2→3) and α(2→6) linkages with other sugars whereas it forms α(2→8) and α(2→9) linkages with other sialic acids (Angata and Varki 2002). These structurally diversified forms of sialic acid can govern biological recognition processes through their critical roles at molecular level.

1.9 Biological roles and functions of Sialic acids

Sialic acids show their presence in many animals and mammals including human body fluids such as blood plasma and breast milk. Because of their negative charge and terminal location, sialic acids are involved in many biological events especially in cellular and molecular recognition processes (Nelson et al., 1995b; Gahmberg and Tolvanen, 1996; Bucior and Burger, 2004; Olofsson and Bergström, 2005; Varki and Gagneux, 2012). Sialic acids serve as good receptors in many pathogenic processes involving bacteria, viruses and parasites (Corfield et al., 1983; Schauer 1985; Wurzer et al., 2002; Vimr et al., 2004; Zeng et al., 2008). Notably, acetylated derivatives of sialic acids participate in many pathogen binding processes (Priyadarzini et al., 2009). As cell surface receptors, the various linkages of sialic acids with other sugars play crucial roles in pathogenic binding processes. Specifically, Neu5Acα(2→3)Gal and Neu5Acα(2→6)Gal (shortly denoted as N23G and N26G respectively) are cell surface receptors which are recognized by various toxins and viruses with different binding specificities. Preferentially, avian influenza viruses recognize sialic acids with α(2→3) linkages while human influenza viruses recognize sialic acids with α(2→6) linkages. But Swine flu viruses can recognize
both $\alpha(2\to3)$ and $\alpha(2\to6)$ linked sialic acids with different binding specificities (Makita and Taniguchi, 1985; Angström et al., 1994; Berman et al., 2000; Brocca et al., 2000; Ishida and Kiso, 2001; Neu et al., 2011; Priyadarzini et al., 2012b). Sialic acids play significant roles in cellular adhesion between cancer cells and endothelial cells and their concentration is connected to the metastatic potential of tumors (Wang, 2005; Malati, 2007). Also sialic acids are potentially used as markers for cancer and cardiovascular diseases (Roseman, 1970; Gopaul and Crook, 2006; Cui et al., 2011; Mitic et al., 2012).

### 1.10 Influenza Virus

Influenza is a significant viral disease caused by the influenza viruses belonging to Orthomyxoviridae family. In humans, the infection targets the columnar epithelial cells of the respiratory tract that cause primary viral pneumonia, which can result in death (Julkunen et al., 2001; Thomas et al., 2006; Steel et al., 2010) Influenza viruses are enveloped, negative stranded and roughly spherical in shape (figure 1.13). The surface of influenza virus is majorly decorated with two different glycoprotein spikes and they are Hemagglutinin (HA) and Neuraminidase (NA). HA is about four times more abundant than NA (Varghese et al., 1983; Nicholson et al., 1998) in the surface. HA is the receptor-binding and membrane fusion glycoprotein of influenza virus since it initiates viral infection process through binding with sialic acid-containing cell surface receptors (Skehel and Wiley, 2000). The NA acts as a receptor-destroying enzyme hence it cleaves the sialic acid from the underlying sugar chain to facilitate the release of viral progeny from the infected host cells and to promote the spread of the infection to
neighbouring cells (Suzuki, 1997). In addition to these two glycoprotein spikes, M2 protein is also present in the virus surface, which functions as an ion channel for the acidification of the interior of the viral particle during viral infection (Pinto et al., 1992; Wang et al., 1994). The central core of the virus, ribonucleoprotein (RNP) is involved in the transcription of viral genes and replication of the viral RNA genome in the nucleus of the infected cells (Cheung and Poon, 2007; Fodor, 2012). Influenza viruses are classified into A, B, and C types according to the antigenic differences between their nucleoproteins and matrix proteins (Cheung and Poon, 2007). In
addition, based on the antigenic variation of the hemagglutinin and neuraminidase, Influenza A viruses are further classified into different subtypes. Till now, 17 subtypes of HA and 11 subtypes of NA have been found (Laver et al., 1984; Fouchier et al., 2005; Tong et al., 2012; Tong et al., 2013) and mostly, waterfowl is a source for all of them (Webster et al., 1992; Olsen et al., 2006; Stallknecht and Brown, 2007). Influenza is also generally called “flu” and H1N1 and H5N1 subtypes are usually known as “Swine flu” and “Bird flu” respectively. Spanish flu 1918 (H1N1), Asian flu 1957 (H2N2), Hong kong flu 1968 (H3N2), Russian flu 1977 (H1N1) and Mexican flu 2009 (H1N1) are the most remarkable epidemics of influenza in the recent times (Kawaoka et al., 1989; Ha et al., 2002; Johnson and Mueller, 2002; Taubenberger and Morens, 2006). In spite of the number of human responsible subtypes of HAs and NAs is high, only H1N1, H2N2, H3N2, H5N1, H7N7 and H9N2 subtypes have been isolated from human (Claas et al., 1998; Subbarao et al., 1998; Yuen et al., 1998; Guan et al., 1999; Lamb and Krug, 2001; Guan et al., 2003; Fouchier et al., 2004) and this signifies that there is a host restriction for influenza viruses. Certain regions of the binding site of HA are highly conserved between subtypes of the influenza virus (Daniels et al., 1984).

1.11 Significant receptors for Influenza virus

The infection process of influenza virus is initiated by hemagglutinin which starts the binding process by recognizing sialic acid receptors protruding from the cell surface of the host cell (Sauter et al., 1989; Eisen et al., 1997; Gambaryan et al., 1997; Martín et al., 1998; Kaverin et al., 2000; Wang et al., 2004; Ghedin et al., 2005; Schwahn et al., 2009). Specifically, hemagglutinin recognizes two significant cell
**Figure 1.14** Influenza virus infections through hemagglutinin binding with cell surface sialic acid receptors

surface sialyldisaccharides, N23G and N26G. Evidently human influenza viruses recognize N26G whereas avian influenza viruses recognize N23G (Baum and Paulson, 1989; Ito et al., 1997; Ryan-Poirier et al., 1998; Ito et al., 2000; Suzuki, 2005; Chua and Chai, 2012; Priyadarzini et al., 2012b). But Swine flu viruses can recognize both the receptors. Figure 1.14 shows the interaction between the hemagglutinin of influenza virus and cell surface sialic acid of human respiratory tract. The binding specificity of influenza hemagglutinins can be altered by making suitable amino acid mutations at the receptor binding site (Rogers et al., 1983; Ledesma et al., 2011).
1.12 Biophysical techniques to study protein–carbohydrate interactions

Molecular level investigations on the structure and conformations of protein–carbohydrate interactions can be carried out through X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy which are the well known and widely used biophysical techniques. X-ray diffraction is a powerful tool in the structure determination of proteins, nucleic acids and carbohydrates. NMR spectroscopy is a powerful and theoretically complex analytical tool in determining the biomolecular structures. Moreover, it is a principal experimental technique to explore the solution structure of biomolecules including protein–carbohydrate complexes. It is not possible to obtain all the structural details of given biomolecule through a single technique, thus a variety of experiments are necessary to elucidate the detailed structure of biomolecule. The inadequacy of experimental techniques demands the use of theoretical techniques in determining and complementing the biomolecular structure. Computational techniques can assist in determining new structures and in structural refinement of existing structures. Molecular dynamics simulations and quantum mechanical calculations are the broadly used theoretical or computational methods to investigate the biomolecular structures including protein–carbohydrate complexes (Priyadarzini, 2012; Selvin, 2013; Parasaruman, 2015; Murugan et al., 2015). Molecular dynamics simulation is run based on Newtonian mechanics and the evolution of the molecular system at particular duration of time in the required biological environment is calculated. The positional trajectories are collected at each time step and by analyzing these emerging trajectories, molecular properties of the molecular system can be understood. In quantum mechanical calculations,
wave functions are used to calculate the state of the molecular systems. Mostly, computational techniques are used in tandem with experimental techniques and provide complementary results (van Gunsteren and Berendsen, 1990; Sharmila and Veluraja, 2004; Veluraja and Margulis, 2005; Kandt et al., 2007; Priyadarzini et al., 2012a; Selvin et al., 2012; Parasuraman et al., 2014; Murugan et al., 2015). The structural coordinates of the biomolecules are available in databases. Three dimensional coordinates of small molecules especially glycans are available in Cambridge Structural Database (CSD). Three Dimensional Structural Database for Sialic acid containing CARbohydrates (3DSDSCAR) offers plausible conformational models for glycans generated through MD simulation trajectories (Veluraja et al., 2010). Similarly, three dimensional structural coordinates of protein–carbohydrate complexes can be downloaded from the various structural data banks such as Protein Data Bank (PDB), SWISS–MODEL, National Centre for Biotechnology Information (NCBI), and Orientations of Proteins in Membrane (OPM).

1.13 Hydrogen Bonds

A hydrogen bond (H–Bond) is the attractive force between the hydrogen attached to an electronegative atom of a molecule and an electronegative atom of a same or different molecule. Simply, hydrogen bond is an attractive force between two electronegative atoms through hydrogen. Commonly, the electronegative atom is oxygen, nitrogen, or fluorine (Vinogradov and Linnell, 1971; Stahl and Jencks, 1986; Zheng and Merz, 1992; Elstner et al., 2001). Generally fluorine involved hydrogen bonds are very strong because fluorine is highly electronegative in nature. Fluorine bonding in protein–carbohydrate interactions have been reported
Symbolically hydrogen bond interaction is denoted as D–H—A, where D is the donor atom and A is the acceptor atom. A hydrogen atom attached to an electronegative atom will play the role of the **hydrogen bond donor** (D) since it donates hydrogen to form hydrogen bond. An electronegative atom such as fluorine, oxygen, or nitrogen will be the **hydrogen bond acceptor** (A), whether it is bonded to a hydrogen atom or not. In hydrogen bonds, the distance between hydrogen atom and acceptor atom is called hydrogen bond length and donor–acceptor angle $\angle$DHA is called hydrogen bond angle (figure 1.15). In biomolecules, Ramachandran criteria states that the distance between two electronegative atoms D (donor) and A (acceptor) is must be 2.6 $\text{Å}$ to 3.2 $\text{Å}$ for a hydrogen bond. For an ideal hydrogen bond, the hydrogen bond angle $\angle$DHA must be 180° but generally it lies between 150° to 180° (Rajagopal and Vishveshwara, 2005; Arunan et al., 2011).

The hydrogen bond attractions can occur between molecules or within a single molecule and called intermolecular and intramolecular interactions respectively. The hydrogen bond is usually stronger than a van der Waals interaction, but weaker than covalent or ionic bonds. This type of bond can occur in water, DNA and proteins.

Intermolecular hydrogen bonding is responsible for the high boiling point of water (100°C) and intramolecular hydrogen bonding is partially responsible for the secondary and tertiary structures of proteins and nucleic acids. The hydrogen bonds examined in anti–parallel $\beta$–sheets of protein secondary structures are linear ($\angle$DHA=180°) whereas the hydrogen bonds in parallel $\beta$–sheets are non–linear ($150° \leq \text{DHA} \leq 180°$) (figure 1.16).
Figure 1.15 (a) Hydrogen bonds in water molecules and (b) Representation of hydrogen bond length and angle $\angle DHA$
Figure 1.16 Hydrogen bonds in protein secondary structures, (a) parallel and (b) anti-parallel β-sheets
Though hydrogen bonds occur in between water molecules, a single water molecule can mediate the hydrogen bonds between active site residues of protein and carbohydrate towards stabilizing certain conformers of protein–carbohydrate complexes (Stahl and Jencks, 1986; Saenger and Jeffrey, 1991; Zheng and Merz, 1992; Jeffrey, 1997). These types of hydrogen bonds are called water mediated hydrogen bonds. Hydrogen bonds significantly govern the stabilization of various conformers of protein–carbohydrate complexes (Quiocho, 1986; Imberty et al., 1991; Weis and Drickamer, 1996; Priyadarzini et al., 2012b; Parasuraman et al., 2014; Parasuraman et al., 2015).

1.14 Objectives of the study

To design the hemagglutinin based inhibitor or to design glycan based drugs against influenza, it is essential to understand the binding specificity of hemagglutinin–sialyldisaccharides complexes. A complete understanding of the binding specificity and conformational features of the above complexes can be obtained by carrying out longer duration molecular dynamics simulation studies and quantum mechanical calculations.

Hence the present research is focused on,

1. To develop a program to generate the input file for Quantum Mechanical calculations from MD simulation trajectories to perform geometry optimization of the protein–carbohydrate complexes, H1–N23G and H1–N26G.
2. To perform Quantum Mechanical Geometry optimization H1–N23G and H1–N26G using Gaussian09 software.

3. To investigate the binding specificity of fluorinated sialyloligosaccharide Neu5Aca(2→3)Gal with influenza hemagglutinin H1 through 100 ns molecular dynamics simulation study.

4. To investigate the binding specificity of fluorinated sialyloligosaccharide Neu5Aca(2→6)Gal with influenza hemagglutinin H1 through 100 ns molecular dynamics simulation study.