S4-1

The Necessity and General Strategy for the Structure Determination of Complex Carbohydrates in Glycoconjugates

Yong-il Park
Division of Biotechnology, The Catholic University of Korea

During the last decade, there has been a vast expansion in our knowledge of the distribution, structure, biochemistry, biosynthesis, and roles of glycoconjugates in nature. The interest over the carbohydrate moieties of glycoconjugates has been increasing since they are known to be involved in diverse array of biological phenomena such as in cell adhesion and recognition, signal transduction, differentiation and carcinogenesis, fertilization, modulation of immune function and enzyme activity, intracellular trafficking and localization of glycoproteins, clearance of glycoproteins in circulation, infectious diseases, and acting as cell surface receptors for lectins, antibodies, toxins, etc. (Lis & Sharon, 1993; Varki, 1993). Also, the effects of different glycoforms on the stability and biological activities of recombinant glycoproteins need much concern in the biotechnological production of pharmacologically useful glycoproteins (Parekh *et al.* 1989).

The carbohydrate moieties of the major classes of glycoconjugates include those of glycoproteins, glycolipids, proteoglycans, and GPI-anchored proteins. Glycosylation is the most common and most versatile modification of proteins, with highly diverse carbohydrate structures ranging from mono- to polysaccharides attached to the protein. There are two major types of vertebrate glycosylation: *O*-glycosylation (thus named *O*-glycan), where the reducing end of, in most animal glycoproteins, the innermost *N*-acetylgalactosamine (GalNAc) is bound to the hydroxyl of a serine or a threonine residue, and *N*-glycosylation (thus named *N*-glycan), where the reducing end of the innermost *N*-acetylglucosamine (GlcNAc) is attached to the amide function of an asparagine in the consensus sequence Asn-X-Ser/The, X being any residue but a proline (Lis & Sharon, 1993; Varki, 1993). In general, *N*-glycan chains are longer than *O*-glycan chains, and nearly all animal *N*-glycans evercharacterized contain the pentamer core oligosaccharide, Manα1-6(Manα1-3)Manβ1-4GlcNac β1-4GlcNAcβ1-Asn. Another type of protein modification by carbohydrates is the attachment of the C-terminal end of hydrophilic protein to the amide function of ethanolamine of anchor oligosaccharide of GPI (glycosylphosphatidylinositol) in plasma membrane. The GPI-anchored proteins can be distinguished from the membrane spanning proteins in that they are released from the membrane by PI-PLC (phosphatidylinositol specific phospholipase C) action (Ferguson, 1991). Proteoglycans are different type of glycoproteins since their sugar chains (GAG, glycosaminoglycan) are much longer (100-200 monosaccharide residues) than the regular *N*- and *O*-linked sugar chains and consist of typical disaccharide repeating units with many anionic residues, such as uronic acids and *O*- and *N*-sulfated sugars (Noonan and Hassel, 1993). Glycolipids, such as gangliosides, are one

of the major group of glycoconjugates in nature; their carbohydrate moieties are generally about 1-10 residues long and are known to be involved in differentiation, signal transduction and binding of infecting bacteria or toxins (Saito, 1993).

The structural diversity of the complex carbohydrates in glycoconjugates

In comparison with genes and proteins, carbohydrates in glycoconjugates, especially in glycoproteins, show considerable diversity in their structure, known as structural microheterogeneity; while protein sequence is completely encoded by the genome, the sequence and structure of the sugar chain only depend on the action of highly specific and precisely located enzymes known as glycosyltransferases and glycosidases. Unlike protein and nucleic acid, each monosaccharide contains several free hydroxyl groups capable of bonding with a second monosaccharide to form up to eight positional and anomeric isomers (α or β). In this way, oligosaccharides, even with the same number and same kinds of monosaccharides, can form glycoforms that differ not only in sequence and chain length, but also in anomery, position of linkages and branching points, resulting in the generation of an enormous variety of glycans with different structures. Also, the structural diversity of glycans originates from the fact that glycosylation is cell- and tissue-dependent and can change with the age and the physiological and pathological states, for example, as it has been reported that, upon transformation, the glycan structures of some surface proteins of normal cells change to multi-antennary oligosaccharides (Kobata, 1996). Moreover, in an individual glycoprotein more than one carbohydrate unit is often present, attached at different positions by either an *N*-linkage, an *O*-linkage or both so they generally present as a highly heterogeneous mix of different species bound to a single protein.

The necessity of structure determination of the complex carbohydrates in glycoconjugates

Structural variation in carbohydrate moieties of the glycoproteins, even in a single protein, sometimes modulate the glycoprotein properties, such as stability, enzymatic activity, antigenicity, solubility, cellular processing, secretion, clearance in circulation, and binding affinity to and specificity for other biomolecules (Varki, 1993). Thus, it is important to know the structure of the carbohydrates present in glycoconjugates for 1) elucidation of more structures and their more precise biological functions in nature, 2) diagnosis of certain cancer and other glyco-related diseases such as congenital disorders of glycosylation (CDGs) (Freeze, 2001), 3) chemical or enzymatic synthesis of structural derivatives for lectin interaction study and development of potent carbohydrate drugs, and 4) biotechnological production of human glycoproteins with therapeutic activity from an appropriate expression system. However, elucidation of the structure of carbohydrate is still challenging and demanding since, in addition to the enormous diversity of carbohydrate structures and the greater technical difficulties, complex carbohydrates can not be amplified nor synthesized in large quantities, which gives researchers great difficulties in dealing with glycoconjugates.

General Strategy for the Structure Determination of Complex Carbohydrates in Glycoconjugates

Methods for complex carbohydrates analysis remain largely non- automated and far more labor

intensive and diverse than those for protein and DNA sequence analysis, where automated sequencers and synthesizers offer analytical power and speed. In most cases so far elucidation of complete structure of carbohydrate moieties of glycoconjugates can be obtained only by combinations of the data obtained from the serial chemical, enzymatic, and instrumental analyses: isolation and purification of glycoconjugates, monosaccharide composition analysis by HPAEC/PAD or reversed phase ODS/fluorescense HPLC system after acid hydrolysis, enzymatic or chemical release and labeling and fractionation of oligosaccharides by chromatographic methods, sequential digestion of glycans using exoglycosidases for the detection of some monosaccharides and their specific linkage analysis, lectin affinity column chromatography for efficient fractionation of different glycoforms, ^{13}C and ^{1}H NMR spectroscopy and mass analysis, methylation analysis for complete linkage analysis (Lennarz and Hart, 1994; Lee, 1990; Fan *et al.*, 1994; Lis & Sharon, 1993). Monosaccharide composition analysis is a good starting point for the structure analysis since it is necessary to choose a strategy for the structural study of the sugar chains of glycoconjugates (Park *et al.*, 1999). Finally, knowing, from the literatures, the typical structures of glycoproteins and ability of glycosylation of individual expression system, such as species of bacteria, yeasts, insects, plants, animals including human cell lines, is also a good strategy to choose appropriate methods for the structural analysis of recombinant glycoproteins, and thereby one can choose or develop a more efficient expression system for industrial production of biologically active glycoproteins with pharmacological properties.

References

1. Fan, J. Q., Y. Namiki, K. Matsuoka, and Y. C. Lee. 1994. Comparison of acid hydrolytic conditions for Asn-linked oligosaccharides. Anal. Biochem. 219:375-378.
2. Ferguson, M. A. J. 1991. Lipid anchors on membrane proteins. Curr. Opin. Strust. Biol. 1:522-529.
3. Freeze, H. H. 2001. Update and perspectives on congenital disorders of glycosylation. Glycobiology. 11:129R-143R.
4. Kobata, A. 1996. Cancer cells and metastasis: The Warren-Glick phenomenon - a molecular basis of tumorigenesis and metastasis. *In* Chap. 3C Glycoproteins and Disease, (*ed.*) Montreuil J., JFG Vliegenthart, H. Schachter. Elsevier, Amsterdam.
5. Lee, Y. C. 1990. High-performance anion-exchange chromatography for carbohydrate analysis. Anal. Biochem. 189:151-162.
6. Lennarz, W. J., and G. W. Hart. (*ed.*) 1994. Guide to techniques in Glycobiology. Methods Enzymol. Vol. 230. Academic Press, INC.
7. Lis, H., and N. Sharon. 1993. Protein glycosylation: structural and functional aspects. Eur. J. Biochem. 218:1-27.
8. Noonan, D. M., and J. R. Hassel. 1993. Proteoglycans of the basement membrane. *In* Molecular and cellular aspects of basement membranes" (*ed.*) D. H. Rohrbach, R. Timpl. Academic Press, New York.

9. Parekh, R. B., R. A. Dwek, J. R. Thomas, et al. 1989. Cell-type- specific and site-specific N-glycosylation of type I and type II human rissue plasminogen activator. Biochemistry. 28:7644-7622.
10. Park, Y. I., H. A. Wood, and Y. C. Lee. 1999. Monosaccharide compositions of *Danaus plexippus* (monarch butterfly) and *Trichoplusia ni* (cabbage looper) egg glycoproteins. Glycoconjugate J. 16:629-638.
11. Saito, M. 1993. Bioactive gangliosides: differentiation inducers for hematopoietic cells and their mechanism(s) of actions. Adv. Lipid Res. 25:303-327.
12. Varki, A. 1993. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 3:97-130.