Glycoinformatics and Glycosciences

Anita Sarkar
The Scripps Research Institute, USA

Serge Pérez
Centre de Recherches sur les Macromolécules Végétales (CERMAV), France

INTRODUCTION

Complex carbohydrates are built for high-density biocoding, which is at par with proteins and nucleic acids and their role and importance is widely being recognized. This can be conceptualized as an extended paradigm of molecular biology in which biological information flows from DNA to RNA and protein. This article describes the role of glycoinformatics in the growth of glycobiology and in the area of structural characterization of glycans, within the area of carbohydrate research.

BACKGROUND

Glycans, both in the form of homopolymers or bound to proteins and lipids, are the most abundant class of biomolecules and are increasingly being implicated in health and environmental issues. Glycosylation is by far the most important post-translational modification (PTM) in terms of the number of proteins modified and the diversity generated. Since glycoproteins, glycolipids and glycan binding proteins (GBPs) also called lectins, are located on the cell surface, many biologically significant events can be attributed to glycan recognition, like those found between mammalian eggs and sperms, between pathogen/parasites to establish the infection process and by the host's immune system in its attempt to combat infection.

GLYCOBIOLOGY TO GLYCOMICS

The term glycobiology was introduced during the 1980s (Rademacher, et al., 1988), when it became apparent that a knowledge of sugar decorations and carbohydrate structures (glycocode) was necessary to fully describe biomolecular functions. Glycobiology was born at the interface of biochemistry, carbohydrate chemistry and molecular biology with the aim of studying the biosynthesis, structure and biological functions of saccharides (sugar chains/glycans). Later, the conceptual term glycome was described in the literature to refer to the complete set of glycan structures synthesized by an organism, at par with proteome and genome (Feizi, 2000). The term glycomics has become common, in analogy to genomics and proteomics (Taniguchi *et al.*, 2001) and emphasizes the holistic view of the total glycan content and its functions in a given organism, cell or tissue. Protein-carbohydrate interactions play crucial roles in the onset, detection, and, potentially, also the prevention of diseases such as cancer, inflammation, diabetes, neurodegenerative disorders, bacterial, viral infections. Such interactions are also involved in the biosynthesis as well as biodegradation of the main raw materials on Earth, thus rapidly becoming an important area of research.

Glycosylation: A Requirement of the Cell

Many observations have led to the conclusions that altered protein profiles of cells and tissues are often a result of an alteration in protein expression rather than a modified gene expression. PTMs in proteins increase manifold the functional diversity of the proteome by modifying the translated protein encoded in the genetic information content, and have a marked influence on all aspects of biological functions and pathogenesis. There is a surprisingly limited number of genes in the entire mammalian genome, than what was expected before the completion of the human genome project. This indicates that PTMs regulating protein functions in the

DOI: 10.4018/978-1-4666-5888-2.ch040

phenotype of cells have a much greater role to play than what was previously acknowledged. This has caused a gradual shift of focus of research from the genetic code to PTMs. Glycosylation is the most extensive and complex form of the protein PTMs providing for the functional diversity to generate multiple phenotypes from a limited genome. The central dogma of cellular biology can be extended as shown in Figure 1.

The variety of information transmission and reception that glycans provide to the genetic code, translated to proteins, though advantageous for the cell, has hindered carbohydrate research.

Glycosylation is highly sensitive to alterations in cellular function, and abnormal glycosylation is indicative and used as a diagnostic technique in diseases. Glycosylation patterns differ among glycoconjugates of different species (driven by evolutionary selection pressures) as well as between different cell types of the same organism. Site-specific protein glycosylation suggests that the 3D structure of the protein has an important role to play in fixing the extent and type of its own glycosylation.

The actual geometry in which the oligosaccharide presents itself to the receptor is important to trigger a biological response (Dwek, 1995). If the same sugar occurring on multiple proteins does not present itself in the correct geometry to the receptor, the receptor cannot distinguish between the 'self' and 'foreign' sugars.

REPRESENTING AND ENCODING CARBOHYDRATE STRUCTURES

Representation in text of the primary structure, or sequence, of complex carbohydrates was first described following the IUPAC-IUBMB terminology in its extended and condensed forms (McNaught, 1997). These forms are used within the carbohydrate community and are adequate for describing complex sugar sequences. Recommendations apply to the description of polysaccharides and glycoproteins (A. McNaught,

Figure 1. Carbohydrates in the scheme of the central dogma of life