Introduction

Zhonghui Sun and Kevin J. Yarema

The term "carbohydrate engineering," has many potential meanings. Strictly speaking, engineering refers to the process of improving a product continuously by standardizing preceding formulations and recombining them to make new and more sophisticated systems. In many respects, based on this definition, Mother Nature herself is the ultimate engineer as she is continually creating novel and increasingly complicated forms of life from sets of "standardized" basic building blocks. Carbohydrates are the perfect example of Nature's engineering efforts as a relatively small handful of basic monosaccharides (nine in humans) are assembled into vastly diverse assemblies of oligo- and polysaccharides that have thousands of structural and regulatory roles. The carbohydrate engineer, upon becoming acquainted with these versatile and incredibly useful molecules, is tempted to jump right in and get to work assisting Nature in designing even more carbohydrate-based structures and adapting them for yet additional applications. But to get started, a number of basic "nuts and bolts" level issues must first be addressed; specifically, one must know what *can* be done, and *how* to do it, by using current and emerging technologies. To address fully these issues and provide the researcher with the optimal set of tools, the term "carbohydrate engineering" is broadly interpreted in this volume.

To some, "carbohydrate engineering" might imply the application of rigorous engineering principles or a technology-based approach to the study or manipulation of carbohydrates. Indeed, engineering and technology have greatly benefited carbohydrate research by providing sophisticated instrumentation, such as that utilized in Chapters 4 to 6[1–3] of this book, which has been indispensable for the recent blossoming of the glycobiology field. Increasingly, "hard-core" engineering-type approaches are being used to complement more empirical, biology-based methods in the study of carbohydrates, as exemplified by the modeling methods outlined in Chapters 7[4] and 8.[5] For others, however, the term "engineering" primarily denotes practicality, especially when used in a biological context. To illustrate, the fields of "genetic engineering" and "tissue engineering" are both highly multi- and interdisciplinary endeavors that eagerly use tools, regardless of if they incorporate any traditional or rigorous engineering aspects, required to get the job done. More specifically, genetic engineering is a mix of microbiology, cell biology, synthetic organic chemistry, biochemistry, and genetics that greatly benefits from technology such as automated DNA synthesizers and sequencers. As an aside, advances in genetic engineering have provided the carbohydrate engineer with powerful tools to manipulate the genes involved in glycosylation. Indeed, many chapters in this book, listed in more detail below in Sections 3 and 4 of the Introduction, exploit recombinant DNA methods to synthesize, modify, and even degrade, complex carbohydrates. Similarly, tissue engineering (the field which seeks to create replacement organs and tissues for worn-out or diseased body parts) exploits and combines advances in cell and molecular biology, polymer chemistry, materials science, developmental biology, and carbohydrates research (especially in the use of polysaccharides as outlined in Chapter 29.[6] This book takes a cue from these fields and broadly defines "carbohydrate engineering" as the study or manipulation of carbohydrates and their biosynthetic processes whatever methods best facilitate the development of practical applications for these complex and versatile molecules.

On a practical level, the carbohydrate engineer is interested in two basic questions; the first is "What use can I find for this carbohydrate?" and the second is "What methods are available to manipulate this carbohydrate?" To address the first issue, namely identifying research, biomedical, or industrial uses for carbohydrates, a basic knowledge of the chemical structure and biophysical properties of these complex molecules is a necessity. Carbohydrates, like proteins, DNAs, and RNAs, are biopolymers built from a basic set of "building blocks," in this case monosaccharides. Unlike proteins and nucleic acids, however, the biosynthesis of carbohydrates is not template-based; moreover they can be assembled into branching structures of greatly increased complexity compared with completely linear macromolecules. These two factors considerably complicate the study and manipulation of carbohydrates, which can be further classified as oligosaccharides and polysaccharides. In general, smaller, branching structures composed of approximately less than 20 ± 15 (definitions are fairly loose here) constituent monosaccharide units are referred to as oligosaccharides; these molecules often play regulatory or signaling roles. Polysaccharides are typically much larger, mostly linear structures that consist of dozens, or even hundreds, of monosaccharide units. A basic review of carbohydrate structures produced in mammalian cells is provided in Chapter 1[7] and specific carbohydrates, or classes of these molecules, are explored in much greater detail in subsequent chapters. Upon knowing the carbohydrates available, efforts continue to exploit these molecules for a host of practical purposes, a few of which are described in this book.

Once a specific application has been determined for a particular carbohydrate, the next question of "What methods are available to isolate or synthesize as well as purify and characterize this complex sugar?" becomes salient. In many cases, a basic understanding of the strategies that Nature uses to create carbohydrates offers valuable assistance in laboratory and biotechnological strategies to manufacture glycans and often Nature's methods can be adapted, sometimes directly and sometimes in roundabout, but highly ingenious, ways by the carbohydrate engineer. For instance, on many occasions, carbohydrates can be directly harvested from biological sources and on other occasions, cells can be employed as factories to produce glycoconjugates or polysaccharides. Cell culture-based production strategies, while they appear simple and lucrative at first because a cell is doing all — or at least much of — the work, nevertheless benefit from a detailed understanding of underlying biosynthetic processes. Metabolic engineers have learned from bitter experience — or in the case of carbohydrate engineers from bittersweet experience — that pathway bottlenecks afflict product quantity and quality. Accordingly, a thorough understanding of the intricacies of the metabolic networks responsible for carbohydrate biosynthesis is a valuable asset in identifying and confronting these problems.

An important benefit of, and indeed often, the primary motivation for the study of the basic glycosylation processes of a cell is to gain clues on how to intervene in diseases where carbohydrate metabolism has gone awry. Therefore the efforts of an engineer seeking to optimize the cellular production of carbohydrates are likely to overlap significantly with the biomedical researcher who studies the role of glycosylation in human disease. In fact, the general idea of an engineer as physician is rapidly gaining currency as evidenced by the number of universities that have established or expanded biomedical engineering departments in the past decade. This book fittingly includes a critical mass of chapters with a biomedical focus, as listed below in Section 6 of the Introduction. Coming back to the topic of practical methods to produce carbohydrates for "real-world" applications, and moving a step farther from Nature's cell-based production methods but still utilizing her tools, glycosylation enzymes are increasingly being used to assemble oligosaccharides with stereo- and regiospecificity that would have proved to be significant challenges to the synthetic chemist. Notwithstanding such challenges, purely synthetic means to produce complex carbohydrates are achieving even more impressive results and efforts are now in progress to make these methods accessible to the nonspecialist.

The definition of "carbohydrate engineering" used in this volume integrates aspects of basic biology, synthetic chemistry, enzymology, complex instrumentation, and sophisticated modeling toward the production of molecules intended for biomedical, research, or industrial applications. Clearly, it is beyond the scope of one volume to address each of these different subdisciplines in depth — indeed, if such an approach was attempted, this book would largely reproduce other specialized books currently available and exclusively devoted to any of the topics of the purification, the physical characterization, the chemical synthesis, or the biosynthetic production of carbohydrates. Instead, the goal of this book is to provide a "big picture" that gives an overview of the diverse tools available to today's carbohydrate engineer. At the same time, however, the goal of each chapter is also to provide detailed technical information designed to be of value to the specialist.

A quick perusal of the table of contents, or a more detailed reading of the chapter descriptions presented below (in Sections I–VII of the Introduction) provides a snapshot of the numerous tools available to the carbohydrate engineer. Moreover, as mentioned, carbohydrate engineering is a truly multi-disciplinary endeavor where techniques and approaches are mixed and matched as needed to increase greatly the diversity of goals capable of being pursued. For organizational purposes, the chapters are first sorted into five sections based on the *methodology* (as listed below). In addition, topics are also categorized based on the *application* (Section VI) or *classification* of sugar (Section VII). In many cases, a single chapter spans (and is discussed in) two or more of these areas and is therefore mentioned multiple times.

I. ISOLATION, PURIFICATION, AND STRUCTURAL CHARACTERIZATION OF CARBOHYDRATES

In general, the abilities to isolate, purify, and characterize complex carbohydrates are necessary prerequisites to the successful manipulation and exploitation of these complex molecules for any practical purpose. The enormous structural diversity of carbohydrates provides a significant obstacle to achieve these ends; for example, a single site of glycosylation on a protein may host one of several dozen potential glycoforms.[8–11] or a single repeating disaccharide unit in a polysaccharide may appear in any one of up to 32 different manifestations after epimerization and sulfation reactions take place. This complexity makes it difficult to separate the many similar, but nevertheless different, forms of naturally occurring (or synthetic) carbohydrates and purify them to homogeneity. In some cases, the separation step can be sidestepped if it is possible to analyze or detect specific structures within an unpurified mixture of sugars as they occur *in situ*. For example, Bovin and co-workers[12] describe a "lectinomics" characterization of cell surface-displayed carbohydrates in Chapter 2, where the objective is to recognize certain epitopes in complex mixtures using lectins or antibodies complementary to a subset of carbohydrate structures (e.g., α-2,6-linked sialic acid or the sialyl Lewis X tetrasaccharide). It can be a daunting, but absolutely crucial task, to use such biological probes to recognize subtle variations in sugar structure, such as the various *O*-acetylated forms of sialic acid recognized as binding sites by human parasites as discussed by Mandal and co-authors[13] in Chapter 3. In other situations, a subset of carbohydrate structures can be altered by metabolic engineering methods to allow the "tagging" and subsequent detection of certain epitopes as discussed by Hinderlich and co-authors[14] in Chapter 13 for sialic acid.

Many times the specialized situations discussed above do not provide an adequate solution to the carbohydrate engineer; instead, the complete physical characterization of a specific carbohydrate of interest is necessary. In these cases, the carbohydrate must be purified to (relative) homogeneity before structural analysis can proceed. In recent years, there has been rapid development of many separation techniques designed to purify complex carbohydrates that include specialized two-dimension (2D) gel electrophoretic[15–19] and HPLC techniques.[20–24] Upon purification, methods must be available to characterize fully an oligosaccharide that are capable of distinguishing between D and L sugars, sequence composition, anomericity, linking positions, and in the case of most oligosaccharides (with the exception of linear polysaccharides), branching structures. Nuclear magnetic resonance (NMR) provides a method for complete structural characterization of carbohydrates of modest size.[25–28] Unfortunately NMR, or complementary methods such as x-ray crystallography,[27,29] require relatively large amounts of material, which is a significant problem considering the trace amounts of many oligosaccharides available after (or even before) purification. Compounding the difficulty of analysis is that, unlike other biomacromolecules such as nucleic acids and proteins, the synthesis of carbohydrates is not template-based. As a consequence, PCR-type amplification methods do not apply, and structural analysis must proceed on the limited quantities of material that can be isolated. Mass spectrometry techniques, in part because of their capacity to analyze minute quantities, are becoming increasingly popular for carbohydrate characterization; these methods often employ instrumentation, such as capillary electrophoresis, which both isolates and characterizes the molecule under test.[30–35] A practical application of sophisticated analytical instrumentation has been to facilitate the diagnosis of congenital disorders of glycosylation (CDG), as discussed by Kościelak in Chapter 4.[1] Another "real-world" situation where carbohydrate characterization is important lies in human milk, where over 900 different oligosaccharides provide a "glycodefense" for the newborn. Clearly, in order to reproduce these sugars commercially, they must first be identified; this task is described by Hueso and colleagues[2] in Chapter 5. To complement the oligosaccharide analysis described in these aforementioned chapters, several chapters also focus on polysaccharide characterization, as exemplified by the thorough discussion of glycosoaminoglycan (GAG) analysis by Sasisekharan and colleagues[3] in Chapter 6.

The conformational analysis of carbohydrates differs in several ways from that of other biomolecules. For instance, many glycans exhibit numerous conformations that coexist in solution at room or physiological temperatures and a conformational analysis of a carbohydrate must address both spatial and temporal properties. Molecular dynamic simulations are well suited for such condensed phase modeling. The interaction between a carbohydrate and a second biological macromolecule, either covalently linked or not, is capable of having a significant impact on either moiety; such interactions are also amenable to computational approaches. Methods to model carbohydrates have undergone rapid development during the past few years with advances in improving force fields, coupling of NMR measurements with molecular dynamics simulations, direct calculation of thermodynamic properties, application of quantum chemical methods on a large scale, and web access to modeling.[36–38] In this book, the basics of structural modeling of polysaccharides are discussed by Stortz in Chapter 7 by using the carrageenan family of compounds as an illustrative example.[4]

II. STUDY OF GLYCOSYLATION IN LIVING CELLS

Cells can be regarded as miniature factories with all the necessary metabolic machinery to produce complex carbohydrates similar to the way they are now exploited for the production of recombinant proteins. However, significant challenges remain before the production of complex sugars in cell culture or fermentation conditions becomes routine (although it should be noted that intense efforts are now in progess, many of which are surveyed in the current book as outlined in the next section). More specifically, although the basic outline of complex carbohydrate biosynthesis is known (as presented in Chapter 1[7]), both the mechanisms that (i) determine the precise carbohydrate structures and (ii) provide high product yields are still not fully explained. Each of these factors is now briefly discussed (below) to illustrate the types of challenges a carbohydrate engineer faces when contemplating the use of a biological system for the manufacture of a sugar-based product.

A continuing puzzle in the biosynthesis of complex carbohydrates is that different classes of glycoproteins, glycolipids, or GAGs are made by enzymes not always unique to each classification group making it unclear how the synthesis of specific structures is maintained. A partial answer lies in the existence of multiple isozymes for several of the transferases that differ in substrate specificity. One enzyme may serve multiple functions, multiple enzymes may serve overlapping functions, or multiple enzymes may form complexes and function together in the synthesis of these macromolecules. Another key factor, alluded to in the previous sentence, is that biosynthetic enzymes form assemblies confined to specific locations within the ER or Golgi. One example of such spatial localization is provided by the stepwise elaboration of *N*-linked glycoproteins by enzymes that habitate either the medial- or