

CHEMISTRY

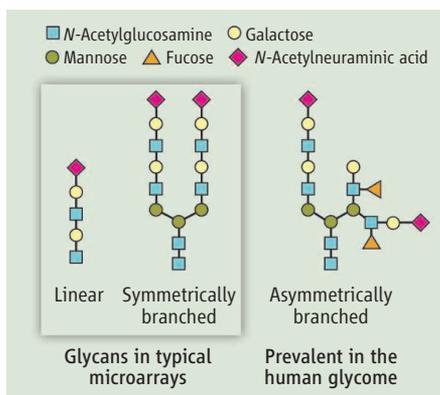
A Path to Complex Carbohydrates

Laura L. Kiessling^{1,2} and Matthew B. Kraft¹

Glycans are carbohydrate-containing compounds that include oligosaccharides as well as the carbohydrate parts of glycoproteins or glycolipids. All cells contain glycans, ranging from those that benefit from the development, metabolism, and immune responses of the host to those involved in cancer metastasis or host-pathogen interactions (1). The oligosaccharide sequence of a glycan determines which proteins it binds to and can thus profoundly influence its physiological activity. The full complement of glycans has not been determined for any cell type, but recent analyses of known mammalian glycans have revealed that there are underlying sequence patterns (2, 3). It remains difficult to assign functions to specific glycan sequences because the relevant oligosaccharides cannot easily be accessed either by synthesis or by isolation from cells (4). On page 379 of this issue, Wang *et al.* report a general and efficient strategy for generating diverse glycans (5).

In the past decade, glycan microarrays have emerged as a critical technology for determining the specificity of carbohydrate-binding proteins (6). Mammalian glycan microarrays are largely populated with saccharides corresponding to terminal fragments of O- and N-linked glycoproteins and glycolipids. The glycans in these arrays are typically linear or symmetrically branched; therefore, the arrays lack the asymmetrically branched structures present in natural glycans (see the first figure).

Emerging evidence suggests that these asymmetrically branched sequences are important recognition motifs. For example, different array platforms for immobilizing oligosaccharides can lead to differences in protein binding specificity (6), suggesting that the density, conformation, or relative orientation of the oligosaccharide units matter. These apparent discrepancies may resolve when oligosaccharides that possess the diversity of substitution patterns found in natural glycans are represented on the glycan arrays. The limitation has been generating this requisite diversity. Glycan sequences are not encoded from a template;



The problems of asymmetry. Current mammalian glycan arrays are populated by linear and symmetrically-branched glycans. Asymmetrically branched *N*-glycans seem to be more prevalent in the human glycome but have been difficult to generate.

consequently, there is no standard means of programming cells to produce specific glycans. It can be difficult to obtain sufficient quantities of pure glycans from biological systems, and synthetic methods have been used to fill this void (7).

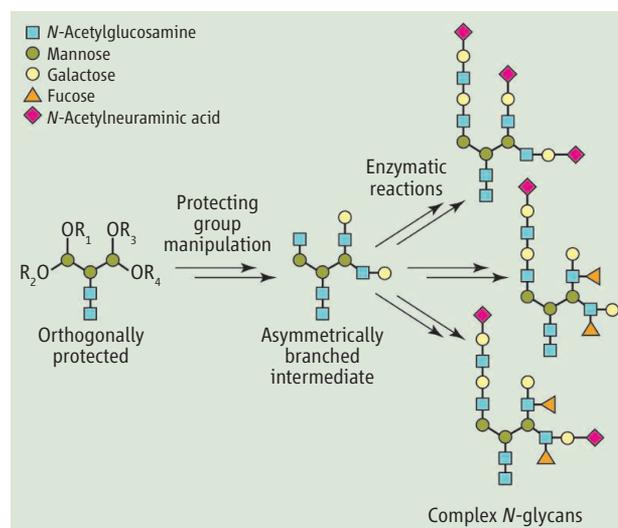
Recent progress in the generation of defined oligosaccharides of specific sequences include the development of robust solid-phase methods, an increased understanding of how protecting groups influence the rates and stereoselectivity of glycosylation, and the advent of new protecting groups that enable excellent chemoselectivity (8, 9). These advances in chemical synthesis have been complemented by the identification and engineering of enzymes that can introduce specific monosaccharide units with exquisite site selectivity and stereoselectivity (10, 11).

Wang *et al.* devise a chemoenzymatic synthesis that takes advantage of progress in chemistry and biochemistry. In their strategy, a single key intermediate serves as a platform for enzyme-mediated diversification to

A general strategy for synthesizing asymmetrically branched carbohydrates will help to elucidate their biological functions.

afford a wide range of branched *N*-glycans. This key intermediate is a protected version of the pentasaccharide core present in all eukaryotic N-linked glycans. By equipping the sites where branching naturally occurs with orthogonal protecting groups, specific sites within the pentasaccharide can be modified independently. Each protecting group is compatible with the synthetic steps needed to build the pentasaccharide, yet each can be removed without modifying any of the others. In this way, each relevant position can be individually elaborated chemically or enzymatically with additional sugar residues to afford asymmetrically branched *N*-glycans.

Another keystone of their approach is the use of glycosyltransferases. These enzymes have long been used for the practical synthesis of oligosaccharides (12). Various naturally occurring and engineered enzymes are now available that mediate selective glycosylation reactions without the need of protecting groups (13, 14). Wang *et al.*'s approach mirrors that of glycan biosynthesis: A core oligosaccharide is first assembled and then customized with a series of enzymatic reactions (see the second figure). For example, the authors first created an asymmetrically branched multiantennary deca-



Chemoenzymatic synthesis of branched *N*-glycans. Wang *et al.* start with an orthogonally protected core pentasaccharide common to all N-linked glycans. Selective protecting group manipulation affords an asymmetrically branched intermediate, which can be further elaborated enzymatically to provide a diverse collection of complex *N*-glycans.

¹Department of Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA. ²Department of Biochemistry, University of Wisconsin–Madison, Madison, WI 53706, USA. E-mail: kiessling@chem.wisc.edu

saccharide with chemical synthesis and then elaborated it with enzymes to provide 13 different complex *N*-glycan structures. They note that 85% of known *N*-glycans are asymmetrically branched and that, in principle, most of these can be accessed by this strategy.

To understand the roles of the core oligosaccharide and the branching structures of *N*-glycans in carbohydrate recognition, Wang *et al.* used a glycan array. They exposed the array to several glycan-binding proteins, including influenza virus hemagglutinins derived from different viral strains. The binding specificities obtained from the arrays containing only linear and symmetrically branched glycans were different than those obtained from arrays with asymmetrically branched glycans. There are mechanisms by which the sequence of each glycan

branch can influence protein recognition: Both branches can interact at an extended binding site, one branch could inhibit binding of another, or a symmetric branched glycan could engage in multivalent interactions (15).

The results reported by Wang *et al.* take us forward in fulfilling the need for well-defined glycans that match the complexity of those found in nature. Their methods provide the means to evaluate how glycan asymmetry influences the ability of glycan-binding proteins to distinguish between different yet related sequences. The study highlights the advantages of combining insights from chemistry and biology to access compounds that would otherwise be difficult to generate. Powerful tools needed to define the functions of specific glycans within the human glycome are now in our purview.

References

1. A. Varki, *Glycobiology* **3**, 97 (1993).
2. D. B. Werz *et al.*, *ACS Chem. Biol.* **2**, 685 (2007).
3. R. D. Cummings, *Mol. Biosyst.* **5**, 1087 (2009).
4. National Academy of Sciences, *Transforming Glycoscience: A Roadmap for the Future* (National Academies Press, Washington, DC, 2012).
5. Z. Wang *et al.*, *Science* **341**, 379 (2013).
6. C. D. Rillahan, J. C. Paulson, *Annu. Rev. Biochem.* **80**, 797 (2011).
7. B. Lepenies, J. Yin, P. H. Seeberger, *Curr. Opin. Chem. Biol.* **14**, 404 (2010).
8. C.-H. Hsu *et al.*, *Angew. Chem. Int. Ed.* **50**, 11872 (2011).
9. M. A. Walczak, J. Hayashida, S. J. Danishefsky, *J. Am. Chem. Soc.* **135**, 4700 (2013).
10. H. A. Chokhawala *et al.*, *ACS Chem. Biol.* **3**, 567 (2008).
11. S. M. Hancocket *et al.*, *Curr. Opin. Chem. Biol.* **10**, 509 (2006).
12. J. C. Paulson *et al.*, *J. Biol. Chem.* **252**, 2363 (1977).
13. L. L. Lairson *et al.*, *Annu. Rev. Biochem.* **77**, 521 (2008).
14. University of Georgia Complex Carbohydrates Research Center, Repository of Glyco-Enzyme Expression Constructs; <http://glycoenzymes.ccrcc.uga.edu/>.
15. L. L. Kiessling, R. A. Splain, *Annu. Rev. Biochem.* **79**, 619 (2010).

10.1126/science.1241788

GENETICS

Genome Mosaicism—One Human, Multiple Genomes

James R. Lupski

With recent advances in genome-wide assays, it is becoming increasingly apparent that a human individual is made up of a population of cells, each with its own “personal” genome. Thus, mosaicism is perhaps much more common within multicellular organisms than our limited genomic assays have detected thus far, and may represent the rule rather than the exception. To what extent does it play a role in normal development and disease?

Chromosomal mosaicism has been recognized clinically for decades, but the application of high-resolution genome-wide analysis tools, such as array comparative genomic hybridization and genomic single-nucleotide polymorphism (SNP) chips has allowed detection of events missed by karyotyping (1–3). Mosaicism for small intragenic copy number variants (CNVs) was detected in 10% of 30 molecularly diagnosed subjects (4), and extensive genomic CNVs have been found in clonal isolates of embryonic stem cells (5). In addition, varying levels of mosaicism have been reported in somatic human tissues, including the skin, brain, and blood, and in

induced pluripotent stem cells (6–10). Therefore, our understanding of the frequency and effects of mosaicism is increasing with the development of ever more sensitive methods for detecting genomic variation.

The Origin of Mosaicism

Mosaicism can arise because of errors that occur during chromosome segregation or DNA replication, leading to chromosome aneuploidy, CNVs, genomic rearrangements, single-nucleotide variation, or repeat expansions and microsatellite instabilities. These mutational processes can occur at any stage of development; in stem cells, differentiating cells, and in terminally differentiated somatic cells.

Both genomic architectural features (e.g., direct and inverted repeats) and DNA sequence characteristics (e.g., CpG dinucleotides) can increase genome instability and susceptibility to mutation. In addition, exogenous sources of DNA damage, such as tobacco smoke and other carcinogens, may lead to somatic mosaicism. Cumulative exposures to exogenous mutagens, as well as ongoing growth of the organism, cell proliferation and renewal, and tissue regeneration, result in accumulation of mutations with age. In addition, early studies suggested that

The cells in our body do not all contain identical genomes, which has implications for health and disease.

nonallelic homologous recombination–predicted inversions (i.e., structural variations) are mosaic and appear to accumulate as the individual ages (11). Somatic mosaicism can also be caused by L1 transposition during embryogenesis (12). Some unbalanced translocations appear to originate postzygotically, apparently arising de novo during embryogenesis in a process that is based on homologous interspersed transposable elements as substrates (13), and other postzygotic mutational events could potentially arise from recombination-restarted replication forks (14).

For both cellular and organismal populations, a balancing act must exist between mutation to generate variation and selection of the variants most fit for that given environment (see the figure). Development itself seems to be a process of strong selective pressure for human genomic integrity. Indeed, single-cell genomic analysis during early development reveals that chromosome instability is common in human cleavage-stage embryos (15), and abnormal chromosome complements can be found in about 70% of 14 normally developing human embryos examined (16); whereas such extensive genomic abnormalities are rare in live-born individuals. Perhaps mutation is tolerated to a

Department of Molecular and Human Genetics, and Department of Pediatrics, Baylor College of Medicine, One Baylor Plaza, Room 604B and Texas Children's Hospital, Houston, TX 77030, USA. E-mail: jlupski@bcm.edu