Chemical Diversity in the Sialic Acids and Related α-Keto Acids: An Evolutionary Perspective

Takashi Angata and Ajit Varki*

Glycobiology Research and Training Center, Departments of Medicine and Cellular and Molecular Medicine, University of California—San Diego, La Jolla, California 92093-0687

Received July 19, 2001

Contents

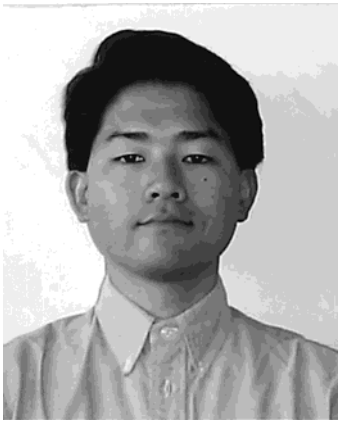
I. Introduction

The glycan chains of vertebrate glycoconjugates are composed mostly of five- and six-carbon sugars. One striking exception is the family of sialic acids (Sia), which are α-keto acids with a nine-carbon backbone. Several other features of sialic acids are unusual. First, they typically occupy the distal end of glycan chains. This "outermost" location makes them suitable for interaction with other cells and with environmental agents, and there are numerous examples of such interactions (described later in this review). Second, sialic acids are subject to a remarkable number of modifications, generating a diverse family of more than 50 structurally distinct molecules (Table 1). Third, unlike the other vertebrate monosaccharides, sialic acids are not ubiquitous in nature and seem to be predominantly found in two distinct branches of the tree of life: in the deuterostome lineage of animals (vertebrates and a few higher invertebrates such as starfish) and in certain types of bacteria. A fourth unusual feature of sialic acids is that they are synthesized via condensation of a neutral six-carbon unit with the three-carbon molecule pyruvate. Finally, sialic acids are unusual in the nature of their high-energy nucleotide sugar

* To whom correspondence should be addressed. Phone: (858) 534−3296. Fax: (858) 534-5611. E-mail: avarki@ucsd.edu.

10.1021/cr000407m CCC: $39.75
Published on Web 01/04/2002

Ajit Varki is currently Professor of Medicine, Professor of Cellular and Molecular Medicine, and Director of the Glycobiology Research and Training Center at the University of California, San Diego (UCSD). He is Executive Director of the textbook *Essentials of Glycobiology* and has served on the Editorial Boards of the several journals, including the *Journal of Biological Chemistry*, *The Journal of Clinical Investigation*, and *Glycobiology*. Dr. Varki is the recipient of a MERIT award from the NIH and an American Cancer Society Faculty Research Award and has served on the Scientific Advisory Boards of the Yerkes Primate Center (Emory University) and the Complex Carbohydrate Research Center (University of Georgia). Significant past appointments include the following: Co-Head, UCSD Division of Hematology-Oncology (1987–1989); President of the Society for Glycobiology (1996); Editor-in-Chief of the *Journal of Clinical Investigation* (1992–1997); the Interim Directorship of the UCSD Cancer Center (1996–1997); and President of the American Society for Clinical Investigation (1998–1999). Dr. Varki's research is currently focused on the sialic acids and their roles in biology, evolution, and disease.



Takashi Angata is currently a postdoctoral fellow at the Glycobiology Research and Training Center, University of California, San Diego. His graduate study at the University of Tokyo focused on the metabolism of KDN, a rare form of sialic acid. Currently he is pursuing the biological functions of the Siglec family of sialic acid-binding lectins. His future research interest includes the roles of sialic acids in vertebrate ontogeny and phylogeny.

donor form. While all the other vertebrate monosaccharides are activated in the form of uridine or guanine dinucleotides, e.g., GDP-Man and -Fuc, UDP-Glc, -Gal, -GlcNAc, -GalNAc, -GlcUA, and -Xyl, sialic acids are activated as cytidine mononucleotides, i.e., CMP-Sia molecules.

Moving beyond vertebrate systems, one finds many more types of five- and six-carbon sugars and a greater variety of linkages and modifications in such sugars (particularly in bacteria). However, there are no other known nine-carbon keto sugars in nature (other than the recently described pseudaminic and legionaminic acids in bacteria, see below). There is

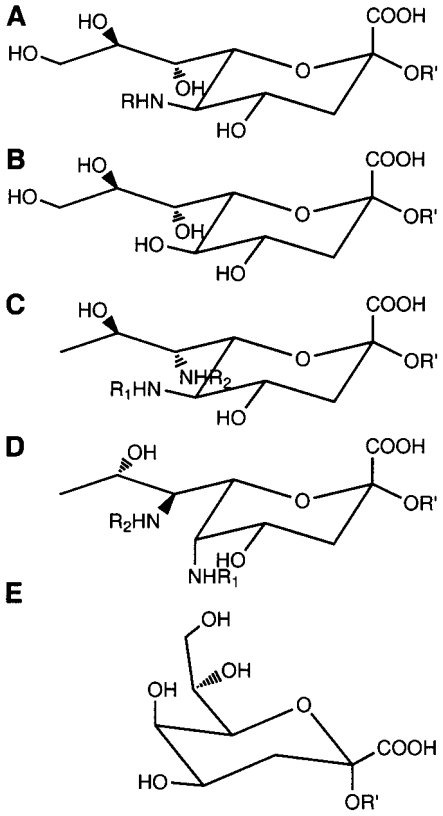


Figure 1. Structures of sialic acids: (A) Neuraminic acid (Neu, R = H), *N*-acetylneuraminic acid (Neu5Ac, R = CH_3-CO−) and *N*-glycolylneuraminic acid (Neu5Gc, R = $HOCH_2$-CO−), (B) 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nonulosonic acid (KDN). Structures of the following sugars are also shown for comparison: (C) 5,7-diamino-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (legionaminic acid) derivative, (D) 5,7-diamino-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid (pseudaminic acid) derivative, and (E) 2-keto-3-deoxy-D-*manno*-octulosonic acid (KDO). Note that the configuration of legionaminic acid was recently amended.[73]

one striking parallel to the sialic acids from both a structural and biosynthetic point of view. The eight-carbon α-keto acid KDO (2-keto-3-deoxy-D-*manno*-octulosonic acid, found in Gram-negative bacteria and in plants; Figure 1E) is formed via condensation of a five-carbon unit (D-arabinose) with pyruvate, and its activated sugar nucleotide form is CMP-KDO. The similarities in the biosynthesis and activation of sialic acids and KDO and the rarity of other similar pathways raise the question of whether the two are evolutionarily related. This review addresses these and other evolutionary issues and provides an overview of sialic acid diversity and its distribution in nature.

The original discovery of sialic acids by Klenk and Blix dates back to the 1930s (discussed in refs 1−3). Although these classic works are beyond the scope of our review, it is worth pointing out that the words "sialic acid" and "neuraminic acid" (the backbone of most sialic acids) both bear the hallmarks of their original discoveries: Blix isolated sialic acid from submaxillary mucin (sialos = saliva in Greek) and Klenk isolated neuraminic acid derivative from brain glycolipids (neuro- + amine + acid). Sialic acids autonomously form a pyranose (six-membered) ring

Table 1. Sialic Acids: Occurrence and Structural Divergence

Compound name	Abbreviation	Occurrence[a]
neuraminic acid	Neu	V[b]
neuraminic acid 1,5-lactam	Neu1,5lactam	V
5-*N*-acetylneuraminic acid	Neu5Ac	V, E, Ps, Pz, F, B
5-*N*-acetyl-4-*O*-acetylneuraminic acid	Neu4,5Ac_2	V
5-*N*-acetyl-7-*O*-acetylneuraminic acid	Neu5,7Ac_2	V, Pz, B
5-*N*-acetyl-8-*O*-acetylneuraminic acid	Neu5,8Ac_2	V, B
5-*N*-acetyl-9-*O*-acetylneuraminic acid	Neu5,9Ac_2	V, E, Pz, F, B
5-*N*-acetyl-4,9-di-*O*-acetylneuraminic acid	Neu4,5,9Ac_3	V
5-*N*-acetyl-7,9-di-*O*-acetylneuraminic acid	Neu5,7,9Ac_3	V, B
5-*N*-acetyl-8,9-di-*O*-acetylneuraminic acid	Neu5,8,9Ac_3	V
5-*N*-acetyl-4,7,9-tri-*O*-acetylneuraminic acid	Neu4,5,7,9Ac_4	V
5-*N*-acetyl-7,8,9-tri-*O*-acetylneuraminic acid	Neu5,7,8,9Ac_4	V
5-*N*-acetyl-4,7,8,9-tetra-*O*-acetylneuraminic acid	Neu4,5,7,8,9Ac_5	V
5-*N*-acetyl-9-*O*-lactylneuraminic acid	Neu5Ac9Lt	V
5-*N*-acetyl-4-*O*-acetyl-9-*O*-lactylneuraminic acid	Neu4,5$Ac_2$9Lt	V
5-*N*-acetyl-7-*O*-acetyl-9-*O*-lactylneuraminic acid	Neu5,7$Ac_2$9Lt	V
5-*N*-acetyl-8-*O*-methylneuraminic acid	Neu5Ac8Me	V, E
5-*N*-acetyl-9-*O*-acetyl-8-*O*-methylneuraminic acid	Neu5,9$Ac_2$8Me	V, E
5-*N*-acetyl-8-*O*-sulfoneuraminic acid	Neu5Ac8S	V, E
5-*N*-acetyl-4-*O*-acetyl-8-*O*-sulfoneuraminic acid	Neu4,5$Ac_2$8S	V, E
5-*N*-acetyl-9-*O*-phosphoneuraminic acid	Neu5Ac9P	V[c,d]
5-*N*-acetyl-2-deoxy-2,3-didehydroneuraminic acid	Neu2en5Ac	V[d]
5-*N*-acetyl-9-*O*-acetyl-2-deoxy-2,3-didehydroneuraminic acid	Neu2en5,9Ac_2	V[d]
5-*N*-acetyl-2-deoxy-2,3-didehydro-9-*O*-lactylneuraminic acid	Neu2en5Ac9Lt	V[d]
5-*N*-acetyl-2,7-anhydroneuraminic acid	Neu2,7an5Ac	V[d]
5-*N*-acetylneuraminic acid 1,7-lactone	Neu5Ac1,7lactone	V
5-*N*-acetyl-9-*O*-acetylneuraminic acid 1,7-lactone	Neu5,9$Ac_2$1,7lactone	V
5-*N*-acetyl-4,9-di-*O*-acetylneuraminic acid 1,7-lactone	Neu4,5,9$Ac_3$1,7lactone	V
5-*N*-glycolylneuraminic acid	Neu5Gc	V, Pz, F
4-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu4Ac5Gc	V
7-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu7Ac5Gc	V
8-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu8Ac5Gc	V
9-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu9Ac5Gc	V, E
4,7-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu4,7$Ac_2$5Gc	V
4,9-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu4,9$Ac_2$5Gc	V
7,9-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu7,9$Ac_2$5Gc	V
8,9-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu8,9$Ac_2$5Gc	V
7,8,9-tri-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu7,8,9$Ac_3$5Gc	V
5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu5Gc9Lt	V
4-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu4Ac5Gc9Lt	V
8-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu8Ac5Gc9Lt	V
4,7-di-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu4,7$Ac_2$5Gc9Lt	V
7,8-di-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu7,8$Ac_2$5Gc9Lt	V
5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu5Gc8Me	E
9-*O*-acetyl-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu9Ac5Gc8Me	E
7,9-di-*O*-acetyl-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu7,9$Ac_2$5Gc8Me	E
5-*N*-glycolyl-8-*O*-sulfoneuraminic acid	Neu5Gc8S	V, E
5-*N*-glycolyl-9-*O*-sulfoneuraminic acid	Neu5Gc9S	E
5-*N*-(*O*-acetyl)glycolylneuraminic acid	Neu5GcAc	V
5-*N*-(*O*-methyl)glycolylneuraminic acid	Neu5GcMe	E
2-deoxy-2,3-didehydro-5-*N*-glycolylneuraminic acid	Neu2en5Gc	V[d]
9-*O*-acetyl-2-deoxy-2,3-didehydro-5-*N*-glycolylneuraminic acid	Neu2en9Ac5Gc	V[d]
2-deoxy-2,3-didehydro-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu2en5Gc9Lt	V[d]
2-deoxy-2,3-didehydro-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu2en5Gc8Me	E[d]
2,7-anhydro-5-*N*-glycolylneuraminic acid	Neu2,7an5Gc	V[d]
2,7-anhydro-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu2,7an5Gc8Me	E[d]
5-*N*-glycolylneuraminic acid 1,7-lactone	Neu5Gc1,7lactone	V
2-keto-3-deoxynononic acid	KDN	V, B
5-*O*-acetyl-2-keto-3-deoxynononic acid	KDN5Ac	V
7-*O*-acetyl-2-keto-3-deoxynononic acid	KDN7Ac	V
9-*O*-acetyl-2-keto-3-deoxynononic acid	KDN9Ac	V
4,5-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN4,5Ac_2	V
4,7-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN4,7Ac_2	V
5,9-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN5,9Ac_2	V
7,9-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN7,9Ac_2	V
8,9-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN8,9Ac_2	V
2-keto-3-deoxy-5-*O*-methylnononic acid	KDN5Me	B
2-keto-3-deoxy-9-*O*-phosphononononic acid	KDN9P	V[c,d]

[a] Abbreviations used: V, vertebrates; E, echinoderms; Ps, protostomes (insects and molluscs); Pz, protozoa; F, fungi; B, bacteria. [b] Present only as bound form. [c] Biosynthetic intermediate. [d] Present only as free form.

in solution via intramolecular hemiketal condensation and adopt a $^2C_5$ chair conformation (Figure 1A,B). In natural glycoconjugates, sialic acids exist only in the α-configuration, except in the high-energy

donor form CMP-sialic acids, where the anomeric carbon is in the β-configuration (α and β configurations refer to the C7 carbon and C1 carboxyl groups being in a trans and cis orientation, respectively).

Sialic acids show remarkable structural diversity in nature, with the family currently comprising over 50 different compounds (Table 1). The original definition of sialic acids was "neuraminic acid (5-amino-3,5-dideoxy-D-*glycero*-D-*galacto*-2-nonulosonic acid, abbreviated as Neu, Figure 1A) and its derivatives".[1,2] This definition was expanded by the discovery of 3-deoxy-D-*glycero*-D-*galacto*-2-nonulosonic acid (or 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid, KDN, Figure 1B),[4] which has a hydroxyl group in place of the amino group at the C5 position. KDN is otherwise very similar to the traditionally defined sialic acids in structure, occurrence (ref 5 and citations therein), and its biosynthetic pathway.[6−8] All other currently recognized sialic acids are biosynthetic derivatives of either *N*-acetylneuraminic acid (Neu5Ac, Figure 1A) or KDN (Figure 1B). *N*-Glycolylneuraminic acid (Neu5Gc, Figure 1A), a major type of sialic acid expressed in deuterostomes, is derived from CMP-Neu5Ac by the addition of an oxygen atom to the *N*-acetyl group, catalyzed by CMP-Neu5Ac hydroxylase.[9] Neuraminic acid with an unsubstituted amino group has been detected in some natural glycoconjugates and is considered to be derived from Neu5Ac via enzymatic deacetylation.[10] Further structural diversity of sialic acids is primarily generated by combinations of the above-mentioned variations at the C5 position with modifications of hydroxyl groups at C4, C7, C8, and C9 by acetate, lactate, sulfate, or phosphate esters or by methyl ethers. Intra- or intermolecular lactonization involving the carboxylate at C1 and intramolecular lactamization between C1 and C5 (in case of Neu) add further variety. Some sialic acids (e.g., 2-deoxy-2,3-didehydro sialic acids) are found only in free form, as expected from the lack of the α-keto group.[11] There are some other sialic acid-like molecules with a nine-carbon backbone found so far only in bacteria, such as derivatives of legionaminic acid (Figure 1C) and pseudaminic acid (Figure 1D). As discussed below, these molecules may be synthesized via very similar biosynthetic pathways to those of sialic acids,[12] raising the possibility of expanding the definition of sialic acids to accommodate them.

Sialic acids are critical in the development of vertebrates, as evidenced by the fact that mouse embryos engineered to lack sialic acid expression die well before they are born (W. Reutter, personal communication). The expression of sialic acids was previously thought to be unique to deuterostomes and to the pathogenic bacteria infecting these animals. However, recent studies have led to the discovery of these sugars in formerly unsuspected organisms, such as certain insects and fungi. Also, the molecular cloning of most of the bacterial and vertebrate enzymes involved in sialic acid biosynthesis (which show significant sequence similarities) has allowed us to seek orthologous sequences in the genomes of other organisms. The available information regarding such genomic sequences has expanded explosively in recent years. Such a "census" survey along with the experimental evidence published in recent years indicate that sialic acids (and/or related molecules) are more widely distributed than previously thought and possibly quite ancient in their origin. In this review we attempt to summarize these recent developments in studies of the structure, biosynthesis, and biology of sialic acids. We also discuss the evolutionary origin of sialic acids, based on existing data and on our phylogenetic analysis of the enzymes involved in their biosynthesis and that of related α-keto acids (see Footnote section). Our analysis leads to a rather unconventional conclusion that the invention of sialic acids may not have happened in the immediate ancestor of deuterostomes (vertebrates, ascidians, and echinoderms). Rather, it may have predated at least the split of deuterostomes and protostomes (arthropods and molluscs). It also may possibly have initially emerged in a lineage of bacteria with later transfer to eukaryotic animals. These and other possible evolutionary scenarios are considered in this review.

Our review mainly focuses on the recent publications concerning sialic acid biology with emphasis on the proteins involved in the biosynthesis and recognition of sialic acids. Many important aspects, such as methodologies in the detection and structural analysis of sialic acids, are not addressed. Readers with interests in these aspects are encouraged to refer to certain classic monographs[13,14] as well as to the recent encyclopedic review elaborated by Schauer and Kamerling.[15] Basic background concerning glycobiology can be found in a recently published textbook.[16] Some technical terms in the fields of genomics and phylogenetics are explained in the Glossary section of this review.

II. Occurrence and Structural Diversity of Sialic Acids in Nature

As shown in Figure 2, all cellular organisms are classified into three domains of life—eukarya (eukaryotes), bacteria (eubacteria), and archaea (archaebacteria). Although there is controversy over the exact evolutionary relationship of these three domains,[17,18] it appears that the original separation of these life forms is quite ancient. Generally speaking, it still stands to be true that the expression of sialic acid is the rule only in deuterostomes and an exception among other life forms. However, the presence of sialic acids has been unequivocally shown in bacteria and some nondeuterostome eukaryotes, and genomic sequence data suggests the presence of sialic acid in an archaeal organism (see below). Figure 2 summarizes the occurrence of sialic acid (based on direct chemical analysis or indirectly based on genomic sequences) in various organisms and their phylogenetic relationship. The generally accepted nomenclature used to denote these molecules is also listed in Table 1. In the following sections we will discuss the presence and variety of sialic acids found in nature, with reference to the enzymes involved in their biosynthesis when appropriate.

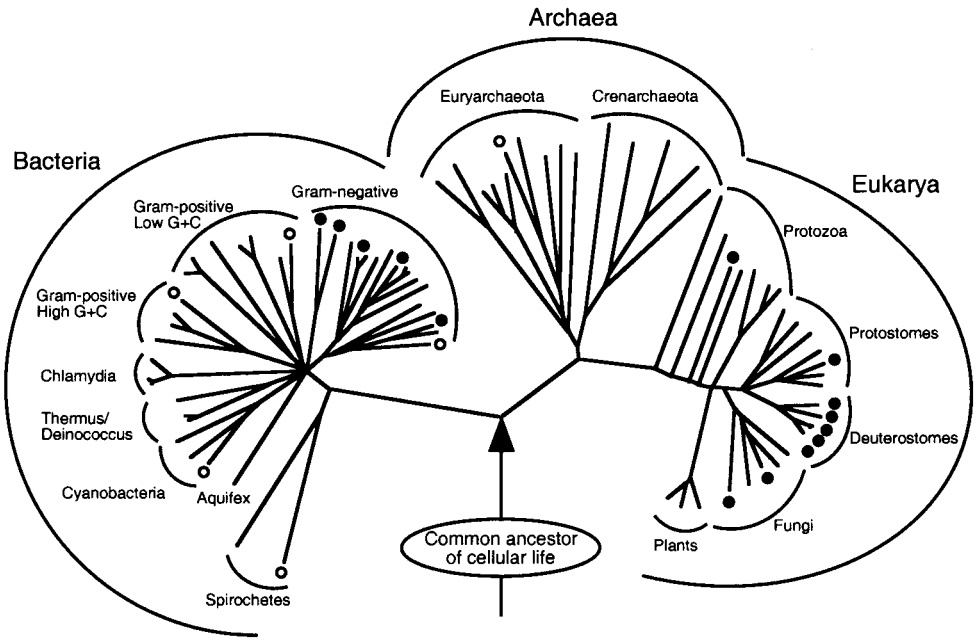


Figure 2. "Universal tree" of cellular organisms and occurrence of sialic acids. The tree was constructed from amino acid sequences of elongation factors (EF-Tu and EF-1α) from various organisms using neighbor-joining method,[206] and the position of the root was estimated by the method described in the literature.[428] Expression of sialic acids (determined by physicochemical methods, e.g., NMR, mass-spectrometry, etc.) is indicated with a closed circle and putative expression (predicted from DNA sequences) with an open circle. There are three examples of apparently complete genomes of eukaryotes that do not show any evidence for the genes involved in sialic acid biosynthesis: *Caenorhabditis elegans* (nematode), *Arabidopsis thaliana* (thale cress, plant), and *Saccharomyces cerevisiae* (baker's yeast, fungus). There are also nine complete genomes of archaea (as of July 2001) that do not show any evidence for such genes.

A. Eukaryotes

1. Animals

Deuterostomes (vertebrates, ascidians, and echinoderms) express four major types of glycoconjugates which are expressed on the cell surface or secreted: glycoproteins, glycolipids, proteoglycans, and glycosylphosphatidylinositol (GPI) anchors. With some exceptions, sialic acids are expressed mostly on the glycan chains of glycoproteins and glycolipids. In most cases they occupy the distal (outermost) end of glycan chains, being commonly linked via an α2−3 linkage to Gal, via an α2−6 linkage to Gal and GalNAc, or via an α2−8 linkage to another sialic acid.[19] Such linkages as α2−3 linkage to GalNAc,[20,21] α2−6 linkage to GlcNAc[22,23] and Glc,[24] α2−4 linkage to Gal[25] and GlcNAc,[26] α2−9 linkage to another sialic acid,[27] and a Neu5Gc oligomer linked via hydroxyl group of *N*-glycolyl[28] also exist but are rare (some are only found, so far, in echinoderms). There are also some examples in echinoderms and amphibians where the sialic acids are found as internal residues, i.e., substituted at one or more hydroxyl groups by another sugar other than sialic acid.[15,29] These "outer" sugars may be considered as additional sialic acid "modifications".

Deuterostomes express the highest structural diversity of sialic acids. All types of *O*-substitutions listed in Table 1 (acetylation, lactylation, sulfation, methylation) can be found in this group. Echinoderms (sea urchins, starfish, etc.) express these modified sialic acids in rather large quantities, which made identification of some of them possible in the past, using less sensitive methods than those available today. Vertebrates were thought to be less "colorful" compared with echinoderms in terms of sialic acid diversity, with mainly *O*-acetylation and occasional lactylation (Table 1). However, a recent report employing a new derivatization method has shown that vertebrate tissues also contain small amounts of rather "unusual" sialic acids, such as 8-*O*-methyl Neu5Ac.[30] Of particular interest among newly discovered sialic acids are Neu 1,5-lactam[31] and Neu5Ac 1,7-lactone.[32] Both of these structures are found in glycoconjugates, and the former may be involved in regulating recognition by selectins (a family of sialic acid recognizing lectins, see section III.B.2), with the latter being specifically recognized by interleukin-4.[31,32] These examples also indicate that subtle modifications of sialic acid molecules can profoundly affect biological interactions.

Protostomes (arthropods and molluscs) seem to express generally similar classes of glycoconjugates as those of deuterostomes. In addition, some organisms (arthropods, such as insects and crustaceans) express an ectoskeleton made of chitin (a polymer of GlcNAc). Despite one earlier report on the presence of sialic acid (Neu5Gc) in a gastropod,[33] protostomes were generally thought to be devoid of this class of molecules. Indeed, the original extensive survey by Warren using the thiobarbituric acid test failed to show a positive response in a wide variety of protostomes examined.[34] Furthermore, with one possible exception,[35] recombinant proteins expressed in insect cells are not reported to have sialic acids.[36,37] However, recent studies have shown that some insects may express Neu5Ac in a stage-specific manner.[38,39] Also, octopuses and squids (belonging to the molluscs)

have recently been shown to express gangliosides (sialic acid-containing glycolipids).[40] Perhaps sialic acids are expressed only in a small number of protostome species. The other possibility is that they are actually common in protostomes but present in very limited quantities and/or restricted to particular tissues/stages. The question also arises as to whether these animals actually synthesize their own sialic acids or simply utilize sialic acid assimilated through the food chain (or synthesized by a commensal organism). In this regard, it is notable that the genome of fruit fly *Drosophila melanogaster* apparently lacks some genes of enzymes necessary for the de novo biosynthesis of sialic acids but contains some others (discussed later).

2. Fungi

There are reports on the presence of sialic acids in some pathogenic fungal cells (reviewed in ref 41), such as *Candida albicans*,[42] *Cryptococcus neoformans*,[43] *Aspergillus fumigatus*,[44] and *Sporothrix schenckii*.[45] The sialic acid species identified so far are Neu5Ac, Neu5Gc, and Neu5,9Ac_2. Although some of these claims are based only on binding of lectins which can recognize sialic acids, others include more solid evidence, such as mass spectrometry.[43,44] We conducted a survey of the available genomic DNA sequences of three fungal species, i.e., *Saccharomyces cerevisiae*, *C. albicans* (http://www-sequence.stanford.edu/group/candida), and *C. neoformans* (http://www-sequence.stanford.edu/group/C.neoformans)—the latter two are still incomplete. This analysis did not reveal any sequences significantly similar to enzymes known to be involved in the biosynthesis, activation, or transfer of sialic acids in bacteria and mammals. However, it is possible that the expression of sialic acids in fungi is strain-specific (as is the case in many bacteria). Alternatively, sialic acids may be acquired from external sources. It also cannot completely be ruled out that these fungi could have independently developed a novel pathway to synthesize and express sialic acids.

3. Plants

With one possible exception, structural studies of natural and recombinant plant glycoproteins[46,47] have shown no evidence of sialic acids. There are a limited number of reports describing sialic acid in plants, of which only one appears conclusive, reporting the presence of Neu5Ac in buckwheat using mass spectrometry.[48] Some earlier reports on sialic acids in plants may be explained by the presence of similar α-keto acids in plants, such as 3-deoxy-D-*arabino*-heptulosonic acid 7-phosphate (DAHP, an intermediate in the schikimic acid pathway of the aromatic amino acid biosynthesis) and KDO (a component of cell wall polysaccharides). Both of these molecules give positive reactions in some traditional colorimetric assays for sialic acids. Our analysis of plant DNA sequences in the GenBank database, including that of thale cress *Arabidopsis thaliana* (the genomic DNA of which is completely sequenced), revealed no evidence for genes known to be involved in the biosynthesis, activation, or transfer of sialic acids.

4. Protozoa

It is known that some protozoa, such as *Trypanosoma cruzi* (which causes Chagas' disease in South America), carry sialic acids.[49] However, these organisms do not synthesize their own sialic acids but rather "borrow" it from their host glycoproteins by the action of a unique enzyme called *trans*-sialidase.[50] The sialic acids are thought to protect these organisms from detection and attack by host immune system.[51] There are some studies claiming the presence of sialic acids (Neu5Ac, Neu5Gc, Neu5,7Ac_2, and Neu5,9Ac_2) on other protozoal cells, such as a slime mold *Dictyostelium discoideum*,[52] a trypanosome *Crithidia fasciculata*,[53] a piroplasmid *Theileria sergenti*,[54] and the amoebae *Entamoeba invadens* and *E. histolytica*.[55,56] Whether these organisms synthesize sialic acids on their own is an unsolved question. There is insufficient genomic data on these organisms to allow a search for genes known to be involved in the biosynthesis, activation, or transfer of sialic acids.

B. Bacteria

Bacteria express various types of glycoconjugates, such as capsular polysaccharides (K-antigens), lipopolysaccharides (O-antigens), S-layer glycoproteins, and peptidoglycans (component of the cell wall). When present, sialic acids are found mostly in capsular polysaccharides and lipopolysaccharides. Unlike the situation in animal glycoconjugates, the sialic acids in these polysaccharides mostly exist as internal residues, either in homopolymers (polysialic acids) linked via α2−8 and/or α2−9 linkages or in repetitive units made of several sugar residues. However, there are also examples of terminal sialic acids in certain bacterial lipooligosaccharides.[57−59] Most of these bacteria express Neu5Ac (with occasional *O*-acetylation, see ref 60 for a comprehensive listing), and a few synthesize KDN and its derivatives.[61,62] Neu5Gc has not, so far, been reported in bacteria.

Most bacteria do not express sialic acids, and even within a bacterial species some strains express sialic acids while others do not (although defining "species" in bacteria is difficult due to frequent exchange of genetic material between species, see refs 63 and 64). Regardless, it is of note that many of the sialic acid-expressing bacteria are causative agents of serious illness in humans and domestic animals and hence became subjects of extensive studies,[65] such as *Escherichia coli* strain K1 and *Neisseria meningitidis* group B,C causing meningitis[66] and *Campylobacter jejuni* causing food-borne gastroenteritis (some strains induce an autoimmune neurodegenerative disorder called Guillain−Barre syndrome).[67] Sialic acids on the cell surface of these bacteria are thought to provide a protective barrier to evade detection and attack by the host's immune system. It was originally assumed that these pathogenic bacteria acquired their sialic acid biosynthetic machinery via horizontal transfer of the genes from vertebrate hosts[68,69]—a rather natural speculation given the conspicuous absence of sialic acids in the organisms evolutionarily situated between bacteria and deuterostomes. How-

ever, there are some bacteria which express sialic acids and are not known to be pathogenic to vertebrates, such as some *Rhodobacter* species[70] and *(Sino)rhizobium meliloti*.[71] In addition, if bacteria acquired a gene from a deuterostome host by lateral transfer in the past, the current bacterial gene should look more similar to the counterpart in a current deuterostome genome (e.g., the one in human genome) than to the equivalent gene in any other genome, e.g., the one in *Drosophila* genome. (Note that since genomic DNA sequences gradually change over time, copies made most recently should look more similar, compared with those made a long time ago). In fact, as discussed later, the bacterial genes involved in sialic acid expression fail to show such an unusually close relation to vertebrate counterparts.

As mentioned earlier, some bacteria express different kinds of 3-deoxy-2-keto-nonulosonic acids, namely, derivatives of legionaminic acid (Figure 1C, originally discovered in *Legionella pneumophila* lipopolysaccharide) and pseudaminic acid (Figure 1D, originally discovered in *Pseudomonas aeruginosa* lipopolysaccharide). Both of these are similar to sialic acids in their structure (reviewed in ref 72; for correct configuration of legionaminic acid, see ref 73) and probably in their biosynthesis.[12] Epimers of legionaminic acids (at C4 or C8-positions) are also reported.[73] Many of the bacteria expressing these sugars are pathogenic in humans; hence, a relation to the virulence is suspected.[72] However, some nonpathogenic bacteria also express these sugars, e.g., *Sinorhizobium fredii* (a symbiotic bacterium of leguminous plants),[74] a marine bacterium *Pseudoaltromonas distincta*,[75] and *Pseudomonas fluorescens*.[76] In addition, these particular sugars have not been identified in vertebrate tissues. Hence, they are likely to be antigenic in vertebrates, which would have an adverse effect on bacterial survival in host tissues. It is therefore likely that these sugars and various modifications have more to do with evading attack by bacteriophages, as suggested for *Sinorhizobium* K-antigens.[77,78]

C. Archaea

Although the structural information on archaeal glycans is limited compared with those of bacteria and eukaryotes, it is known that they do express glycoproteins, glycolipids, and polysaccharides. There have been no published studies concerning the presence of sialic acids in archaea. However, we noted that the genome of a hyperthermophilic archaea, *Methanococcus jannaschii*, contains putative genes similar to Neu5Ac synthetase and CMP-Neu5Ac synthetase in a gene cluster, suggesting that the organism may express sialic acids or a similar molecule. Definite chemical identification of such a compound is awaited. However, the genomes of nine other archaeal species (whose complete genomes were sequenced) do not show evidence for similar enzyme genes.

III. Proteins That Recognize Sialic Acids

Proteins which recognize sugars (excluding antibodies) are collectively called lectins. There are numerous lectins in nature which can recognize sialic acids (Tables 2−4). These are briefly reviewed here to place into perspective their likely role in driving the evolution of chemical diversity in the sialic acids. For convenience, we have grouped them into three categories: vertebrate pathogen lectins, vertebrate endogenous lectins, and lectins from other sources (plants, protostomes, etc.). Members of the third group have great importance as tools for sialic acid research. However, their natural functions are not clearly understood (some are thought to be involved in antibacterial self-defense[79,80]). Thus, this group is not discussed in detail.

A. Vertebrate Pathogen Lectins

Various pathogens (viruses, bacteria, and protozoa) express lectins that can recognize sialic acids. Some pathogens use these lectins for recognition and entry into the host cells, while others express soluble lectins with a sialic acid-binding property. In some cases, these pathogens use not only sialic acids but also some other cell surface molecules for attachment, and different strains show different degrees of dependence toward a particular type of such cellular "receptors".

1. Viruses

Viral sialic acid-recognizing lectins are usually capable of agglutinating red blood cells (hemocytes in old terms) and are thus traditionally called hemagglutinins rather than lectins. Many viruses utilize sialic acids to facilitate attachment to host cells (see listing in Table 2), although their degree of dependence on sialic acids for this purpose varies. Hemagglutinins of influenza viruses (A, B, and C), Newcastle disease virus (NDV, an avian pathogen), mouse polyoma virus, Sendai virus (a rodent pathogen), mouse hepatitis virus, and some others have been isolated and shown to bind sialic acids. Some of these hemagglutinins (from influenza A, C, NDV, and polyoma viruses) have been crystallized and their three-dimensional structure resolved by X-ray crystallography.[81−85] In many other instances (e.g., adenoviruses and picornaviruses), the exact molecular nature of the viral sialic acid-recognizing lectins has not even been identified.

Perhaps the most well-known and well-studied molecule of this kind is the influenza A virus hemagglutinin. Influenza viruses show near obligatory dependence on the host cell surface sialic acids for infection. Sialic acid binding preferences of influenza A hemagglutinins isolated from different host species are correlated with the type of sialic acids expressed on the host cells[86] and even with difference in the sialic acid linkage (α2−3/α2−6). Studies have shown that the natural reservoir of influenza viruses is in various species of wild water fowl.[87] Through a series of cross-species infections involving domestic animals (ducks and pigs) and adaptations to the new types of sialic acids encountered in the new hosts, and/or reassortment of genomic fragments from human and bird influenza viruses in pigs, which are susceptible to both types of viruses,[88] influenza viruses adapt to infect humans and cause seasonal epidemics.

Table 2. Vertebrate Pathogens/Commensal Organisms Which Recognize Sialic Acids[a]

Pathogens (lectin protein, if known)	Specificity[b]	Ref
	Viruses	
Orthomyxoviridae		
Influenza virus A (H)[c] (Fowl plague virus = influenza A)	variable, depending on host and/or strains	86,252,253
Influenza virus B (H)	Neu5Acα2−6Gal	254
Influenza virus C (HE)	Neu5,9Ac_2	89,255
Paramyxoviridae		
Newcastle disease virus (HN)	Neu5Ac/Neu5Gcα2−3Galβ1−4Glc(NAc)	256,257
Sendai virus (HN)	Neu5Acα2−3Galβ1−3GalNAc/4GlcNAc	256,258
Human parainfluenza virus type 1 (HN)	Neu5Acα2−3Galβ1−4GlcNAc	96
Human parainfluenza virus type 3 (HN)	Neu5Ac/Neu5Gcα2−3/6Galβ1−4GlcNAc	96
Porcine rubulavirus LPM (HN)	Siaα2−3Gal	259
Reoviridae		
Reovirus type 3 (σ1)	Sia*	260,261
Porcine rotavirus group A OSU	Neu5Gc/Neu5Acα2−3Galβ1−4Glc (Gc>Ac)	262
Porcine rotavirus group C AmC-1	Sia*	263
Human rotavirus KUN, MO	GM1	264
Human rotavirus Wa	Sia*	265
Simian rotavirus RRV (VP4)	Sia*	266
Simian rotavirus SA11	Neu5Gcα2−3Gal	267
Bovine rotavirus NCDV	Neu5Gcα2−3Gal	267
Bluetongue virus	Neu5Ac, Neu5Gc	268
Coronaviridae		
Bovine coronavirus (S protein)	Neu5,9Ac_2α2−3Gal ≥ Neu5,9Ac_2α2−6Gal	269,270
Human coronavirus OC43 (S protein)	Neu5,9Ac_2α2−6Gal ≥ Neu5,9Ac_2α2−3Gal	269,270
Porcine hemagglutinating encephalomyelitis virus	Neu5,9Ac_2	271
Porcine transmissible gastroenteritis coronavirus (S protein)	Neu5Gcα2−3	272
Avian infectious bronchitis coronavirus	Neu5Acα2−3	273
Mouse hepatitis virus (HE)	Neu4,5Ac_2	95
Adenoviridae		
Adenovirus type 37	Siaα2−3*	274
Adenovirus type 8, 19a	Sia*	275
Picornaviridae		
Bovine enterovirus 261	Sia*	276
Human enterovius type 70	Sia	277
Theiler's encephalomyelitis virus DA	Sia	278
Human rhinovirus type 87	Sia*	279
Encephalomyocarditis virus	Sia	280,281
Mengo encephalomyocarditis virus	Sia*	282
Parvoviridae		
Adeno-associated virus type 4	Neu5Acα2−3Gal	283
Adeno-associated virus type 5	Neu5Acα2−3Gal, Neu5Acα2−6Gal	283,284
Canine parvovirus	Sia*	285
Bovine parvovirus	Sia*	286
Polyomaviridae		
Mouse polyoma virus (VP1)	Neu5Acα2−3Galβ1−3GalNAc	287,288
Polyoma virus JC	Siaα2−6	289
Human polyoma virus BK	Sia	290
Papillomaviridae		
Monkey B-lymphotropic papovavirus	Sia	291,292
Herpesviridae		
Human cytomegalovirus	Neu5Ac > Neu5Gc	293
Mouse cytomegalovirus	Neu5Ac (virulence-associated)	294
Rhabdoviridae		
Rabies virus	Sia*	295
Vesicular stomatitis virus	Sia*	296
Hepdnaviridae		
Hepatitis B virus (small S protein)	Neu5Ac	297
	Bacteria—adhesins or whole cells	
Proteobacteria (Gram-negative)		
E. coli (SfaI, II; common subunit SfaS)	Neu5Gcα2−3Gal, Neu5Acα2−8Neu5Ac	298,299
E. coli (K99 fimbriae)	Neu5Gcα2−3Galβ1−4Glc	104,105
E. coli (F41 fimbriae)	Sia	300
E. coli (CFAI)	Sia*	301,302
E. coli (CFAII; CS2 protein)	Sia (Neu5Gc > Neu5Ac)	303
Helicobacter pylori	Neu5Acα2−3Gal	304,305
Neisseria subflava (Sia-1 adhesin)	Neu5Acα2−3Galβ1−4Glc	306
Pseudomonas aeruginosa	sialyl Le^x	307
	Siaα2−6	308
Haemophilus influenzae	GD2, GD1a, GD1b	309,310
Haemophilus influenzae (HMW1 agglutinin)	Siaα2−3	311
Bordetella bronchiseptica	Neu5Ac	312
Bordetella avium	GD1a, GT1b	313
Pasteurella haemolytica (adhesin)	GlcNAc > Neu5Ac	99

Table 2 (Continued)

Pathogens (lectin protein, if known)	Specificity[b]	Ref
	Bacteria—adhesins or whole cells	
Moraxella catarrhalis	GM2	314
Actinobacillus actinomycetemcomitans	Sia*	315
Gram-positive bacteria		
Streptococcus sanguis (SSP-5)	Neu5Acα2−3Galβ1−3GalNAc	316,317
Streptococcus mitis	Neu5Acα2−3Galβ1−3GalNAc	317
Streptococcus gordonii	Neu5Acα2−3	318
Streptococcus mutans (PAc protein)	Siaα2−6	319
Streptococcus suis	Neu5Acα2−3(Galβ1−4GlcNAcβ1−3)n	320
Streptococcus pneumoniae (CpbA)	Siaα2−6Galβ1−4Glc?	321
Staphylococcus aureus	Sia, sulfated sugar?	322
Ureaplasma urealyticum	Sia*	323
CFB/green sulfur bacteria		
Bacteroides fragilis	Neu5Ac?***,[d]	324
Spirochaeta		
Treponema denticola	Neu5Ac?***,[d]	325
Treponema pallidum	Neu5Ac***	326
Mycoplasma		
Mycoplasma pneumoniae (hemagglutinin)	Neu5Acα2−3(Galβ1−4GlcNAcβ1−3)n	327
Mycoplasma bovis	Sia, sulfated sugar?**	328
Mycoplasma gallisepticum	Sia*	329
	Bacteria—toxins	
Clostridium tetani (tetanus toxin)	GD1b, GT1b	330
Clostridium botulinum (neurotoxin, type A−F)	GQ1b, GT1b, GD1a	331−333
Clostridium perfringens (δ toxin)	GM2	334
Clostridium perfringens (ϵ toxin)	Sia*	335
Vibrio cholera (cholera toxin)	GM1	330,336
Vibrio mimicus (hemolysin)	GD1a, GT1b	337
Vibrio parahemolyticus (thermostable direct hemolysin)	GT1	338
E. coli (heat-labile enterotoxin)	GM1	330
E. coli (heat-stable enterotoxin b)	sulfatide, gangliosides?	100
Bordetella pertussis (pertussis toxin)	GD1a; Neu5Acα2−6Galβ1−4GlcNAc	339,340
Staphylococcus aureus (α toxin)	Neu5Acα2−3Galβ1−4GlcNAcβ1−3Galβ1−4Glc	341
Staphylococcus aureus (γ hemolysin, leucocidin)	GM1	342
	Protozoa	
Plasmodium falciparum (EBA-175)	Neu5Acα2−3Gal	108
Plasmodium falciparum (BAEBL)	Sia*	112
Tritrichomonas mobilensis (TML)	Neu5Ac > Neu5Gc***	343
Tritrichomonas foetus (TFL)	Neu5Ac ≈ Neu5Gc***	113
Tritrichomonas suis	Sia?***,[d]	344
Entamoeba histolytica (toxin)	Sia	345
Toxoplasma gondii	Sia*	346
	Fungi	
Aspergillus fumigatus	Sia	347
Penicillium marneffei	Neu5Ac***	348
Chrysosporium keratinophilum	Neu5Ac***	349
Anixiopsis stercoraria (*Aphanoascus fulvescens*)	Neu5Ac***	349
Microsporum sp. (7 species)	Sia	350
Trichophyton sp. (5 species)	Sia	350
Epidermophyton floccosum	Sia	350

[a] The list compiled here is deliberately inclusive, considering all reports that suggest microbial recognition of sialic acids, including in vitro experiments in various direct and indirect formats. In many cases, the biological relevance of sialic acid binding to in vivo infection process has not been proven. In some instances sialic acids only represent one of several factors involved in initial interactions of microbes with vertebrate cells. Also, due to strain-to-strain difference in cellular "receptor" preference, the fact that one strain uses sialic acid for cell adhesion does not necessarily mean that the other strains of the same microbe also utilize sialic acids. For example, among 102 types of human rhinoviruses tested, type 87 is the only one which requires sialic acid as cell surface receptor.[279] [b] The asterisks (*) indicate the criteria used in the original literature to claim sialic acid recognition by the pathogen (lectin): *, sialidase treatment of host cells/erythrocytes reduce binding/hemagglutination by the pathogen (lectin); **, asialo-counterpart of the glycoconjugate (used as competitive inhibitor) is less effective than intact one in inhibiting binding/hemagglutination by the pathogen (lectin); ***, Neu5Ac (or other mono/oligosaccharide) inhibits binding/hemagglutination by the pathogen (lectin). No asterisks are attached when more than one criteria was used to prove sialic acid recognition or more detailed analysis (e.g., TLC-overlay) was employed to analyze binding specificity. Sia = sialic acid, type not specified. In cases where the specificity is indicated only as Sia, binding of pathogens to sialic acids were determined by criteria * and/or ** as above. In cases in which the evidence is even less robust, we list the item with a question mark (?). [c] H, hemagglutinin; HE, hemagglutinin-esterase; HN, hemagglutinin-neuraminidase. [d] Sialic acid significantly inhibits hemagglutination/binding by pathogen (lectin), but sialidase treatment of erythrocytes/host cells does not inhibit or even enhance hemagglutination/binding.

Among other related molecules, the influenza C virus hemagglutinin-esterase is notable for being specific for 9-*O*-acetylated sialic acids.[89] It not only binds to Neu5,9Ac$_2$, but also has an enzyme activity to destroy the 9-*O*-acetyl ester.[90] This protein has proven useful as a probe for studying 9-*O*-acetylated

Table 3. Vertebrate Lectins Which Recognize Sialic Acids

Names (synonyms)	Expression (source)[a]	Binding specificity	Ref
	Selectins		
E-selectin (ELAM-1, CD62E)	activated endothelium	sialyl Le^x, sialyl Le^a	119,351
L-selectin (MEL14 antigen, CD62L)	leucocyte	6′-sulfo sialyl Le^x, heparansufate	119,351
P-selectin (GMP-140, PADGEM, CD62P)	activated endothelium, platelet	sialyl Le^x, sialyl Le^a, heparan sulfat[e]	119,351
	Siglecs		
Siglec-1 (sialoadhesin)	macrophage	Neu5Acα2−3Gal > Neu5Acα2−6Gal	149,352
Siglec-2 (CD22)	B-cell	Siaα2−6Gal	353,354
Siglec-3 (CD33)	myeloid precursor, Mo[b]	Siaα2−6Gal ≥ Siaα2−3Gal	149,355
Siglec-4 (MAG)	glial cells	Neu5Acα2−3Gal on complex gangliosides	356,357
Siglec-5	Mo, Gr	Siaα2−6Gal ≈ Siaα2−3Gal, Neu5Acα2−8	149,358
Siglec-6 (OB-BP1)	placenta, B-cell	Siaα2−6GalNAc	359
Siglec-7 (AIRM-1)	NK cells, Mo, Gr	Siaα2−6Gal ≈ Siaα2−3Gal (3-Ig isoform)	360
		Siaα2−6Gal (2-Ig isoform)	361
Siglec-8	eosinophil, basophil	Siaα2−3Gal ≥ Siaα2−6Gal	362
Siglec-9	Mo, Gr	Siaα2−6Gal ≈ Siaα2−3Gal	142,363
Siglec-10	B-cell	Siaα2−6Gal ≈ Siaα2−3Gal	364
Siglec-11	?	Neu5Acα2−8Neu5Ac	unpublished
	Others		
Complement factor H	blood	Sia; C7−C9 side chain is a part of epitope	365,366
CD83	dendritic cell	Sia	143
L1	mouse neuron	Neu5Acα2−3 (on CD24)	144
Interleukin-1α	blood	Neu5Acα2−3Galβ1−4GlcNAc, biantennary	32
Interleukin-1β	blood	Neu5Acα2−3Galβ1-Cer (GM4)	32
Interleukin-2	blood	GD1b	367
Interleukin-4	blood	Neu5Ac1,7lactone	32
Interleukin-7	blood	Neu5Acα2−6GalNAc	32
Laminin	extracellular matrix	Neu5Acα2−3Galβ1−4GlcNAc (α2−3>α2−6)	145
Sialic acid-binding protein	endometrium	Neu5Gc > Neu5Ac	368
Sarcolectin	placenta	Neu5Ac, Neu5Gc	147
Calreticulin	ovine placenta	Neu5Gc > Neu5Ac, prefer *O*-acetyl	146
Sialic acid-binding protein	rat uterus	Sia	369
Calcyclin	bovine heart	Neu5Gc	148
Sialic acid-binding protein	frog egg	Sia	370
Ganglioside binding protein*	rat brain myelin	gangliosides (GT1b, GQ1b, GD1b)	371,372
Hemagglutinin*	rat brain	Neu5Ac, Neu5Gc	373

[a] Expression (source) is in human, unless otherwise stated. [b] Mo, monocyte; Gr, granulocyte. [c] Most of the proteins listed here are cloned and/or purified to homogeniety, except for the entities marked with an asterisk (*).

sialic acids.[91−94] Mouse hepatitis virus also has a hemagglutinin-esterase, which is specific to sialic acids substituted by *O*-acetyl group at the C4 position (Neu4,5Ac$_2$).[95] Another prominent example of such bifunctional (receptor−recognition/destruction) proteins are the hemagglutinin−neuraminidases of NDV[84] and parainfluenza viruses.[96] At first glance it appears irrational that a bifunctional protein carries two "contradictory" functions. However, in the case of influenza A virus which carry separate hemagglutinin and neuraminidase proteins, the balance between these two functions appears to be important for efficient viral replication, which involves numerous cycles of host cell attachment−internalization, amplification, and assembly−and detachment from the host cell.[97,98]

2. Bacteria

Some bacterial pathogens interact with host cells in a sialic acid-dependent manner. The involved lectins are attached to the bacterial surface and are typically called adhesins. The identity of the recognition molecule is uncertain in many cases. There are also soluble bacterial lectins, which are typically toxins. Although many have high specificity toward sialic acids, some are not as specific, e.g., recognizing also GlcNAc, as with the *Pasteurella haemolytica* adhesin,[99] or sulfated sugars, as with the *E. coli* heat-stable enterotoxin b.[100]

Bacterial adhesins are often expressed in a strain-specific way and can influence the range of tissues the strain can infect or colonize.[101] Such examples include *Helicobacter pylori*[102,103] and *E. coli* strain K99.[104,105] The former is an etiological agent of peptic ulcers in humans, and the latter causes lethal dysentery among piglets and calves. *H. pylori* express two different adhesins (in an environment-dependent manner) which can recognize sialic acids.[102,103] Despite detailed analyses on ligand structural preference of these adhesins, their exact role in the establishment of infection is not yet understood. The adhesins expressed by *E. coli* K99 strain shows high specificity toward Neu5Gcα2−3Galβ1-4Glc structure on glycolipids, which is abundantly expressed in the gastrointestinal tract of piglets.[106]

Cholera toxin, produced by *Vibrio cholerae*, is a classic example of soluble sialic acid-binding lectins.[107] It is composed of five B subunits and an A subunit. The B subunits show specific binding to a sialylated glycolipid (ganglioside GM1), delivering the A subunit to the cytosol. This in turn causes overactivation of an intracellular signaling pathway (adenylate cyclase, producing cyclic AMP) in gastrointestinal epithelial cells, causing severe diarrhea and

Table 4. Lectins from Invertebrate Animals, Plants and Fungi Which Recognize Sialic Acids

Source[a]	Specificity[b]	Ref
	Mollusca (molluscs)	
Gastropoda (gastropods)		
Limax flavus (yellow garden slug)	Neu5Ac > Neu5Gc	374
Achatina fulica (giant African snail)	Neu5,9Ac_2	375
Cepaea hortensis (white-lip gardensnail)	Neu5,9Ac_2	376
Pila globosa (apple snail)	Neu5Gc	377
Biomphalaria glabrata (bloodfluke planorb)	Neu5Ac[c]	378
Bivalvia (bivalves)		
Anadara granosa (clam/cockle)	Neu5Gc	379
Modiolus modiolus (northern horsemussel)	Neu5Ac (heterogeneous)	380
Crassostrea gigas (Pacific oyster)	Neu5Ac	80,381
Mytilus edulis (blue mussel)	Neu5Ac	382
	Arthropoda (arthropods)	
Chelicerata		
Limulus polyphemus (Atlantic horseshoe crab)	Neu5Ac, Neu5Gc	79,383
Carcinoscorpius rotundicauda (Indian horseshoe crab)	Neu5Gc, Neu5Acα2−6GalNAc-ol	384,385
Tachypleus tridentatus (Japanese horseshoe crab)	Neu5Ac	386-388
Tachypleus gigas (Chinese horseshoe crab)	Sia***	389
Androctonus australis (Sahara scorpion)	sialyllactose, Neu5Ac, Neu5Gc***,[c]	390
Centruroides sculpturatus (bark scorpion)	Neu5Ac, Neu5Gc***	391
Mastigoproctus giganteus (whip scorpion)	Neu5Ac, HexNAc***	392
Heterometrus granulomanus (Indian scorpion)	sialyllactose, Neu5Ac, Neu5Gc***	393
Paruroctonus mesaensis (scorpion)	Sia	394
Vaejovis spinigerus (scorpion)	Neu5Ac, KDO, HexNAc***	395
Hadrurus arizonensis (giant desert hairy scorpion)	Neu5Ac, Neu5Gc > KDO	396
Aphonopelma chalcodes (spider)	Sia, GlcNAc, KDO***	397
Ixodes ricinus (castor bean tick)	Sia, GlcNAc, Gal?	398
Ornithodoros moubata (soft tick)	Neu5Ac, HexNAc	399
Crustacea (crustaceans)		
Cancer antennarius (Pacific rock crab)	Neu5,9Ac_2, Neu4,5Ac_2	400
Birgus latro (coconut crab)	Neu5Ac***,[c]	401
Scylla serrata (mud crab)	Neu5Gc	402
Homarus americanus (American robster)	Neu5Ac, HexNAc	403
Pacifastacus leniusculus (signal crayfish)	Sia?	404
Penaeus monodon (black tiger shrimp)	Neu5Ac, HexNAc	405
Macrobrachium rosenbergii (giant freshwater prawn)	Neu5Ac, Neu5,9Ac_2, HexNAc	406
Tracheata		
Allomyrina dichotoma (beetle)	(Neu5Acα2−6)Galβ1−4GlcNAc	407
Teleogryllus commodus (black field cricket)	Neu5Ac, HexNAc***	408
Hyalophora cecropia (cecropia moth)	Neu5Ac?***	409
	Urochordata (tunicates)	
Halocynthia pyriformis (sea squirt)	Neu5Ac, Neu5Gc	410
Styela plicata (sea squirt)	Neu5Ac, sialyllactose	410
	Viridiplantae (plants)	
Sambucus nigra (European elderberry)	Siaα2−6Gal/GalNAc	411,412
Sambucus sieboldiana (Japanese elderberry)	Siaα2−6Gal/GalNAc	413,414
Sambucus canadensis (North American elderberry)	Siaα2−6Gal/GalNAc	413
Maackia amurensis (Amur maackia)		
leucoagglutinin (MAL)	Neu5Ac/Neu5Gcα2−3Galβ1−4GlcNAc	415,416
hemagglutinin (MAH)	Neu5Acα2−3Galβ1−4[Neu5Acα2−6]GlcNAc	416,417
Trichosanthes japonica (Mongolian snake gourd)	Neu5Acα2−6(or HSO_3−6)Galb1−4GlcNAc	418
(*Trichosanthes kirilowii var. japonica*)		
Saraca indica (Ashoka)	Neu5Acα2−3/6Galβ1−4GlcNAc**	419
Morus alba (white mulberry)	Neu5Gc***	420
Lactuca scariole (*L.serriola*, prickly lettuce)	Sia?**	421
Viscum album (European mistletoe)	Sia slightly enhances binding	422,423
Triticum aestivum (bread wheat)	GlcNAc > Neu5Ac	424
	Fungi (mushrooms)	
Hericium erinaceum	Neu5Gc, Neu5Ac	425
Agrocybe cylindracea	Neu5Acα2−3Galβ1−3GalNAc/GlcNAc	426
Polyporus squamosus	Neu5Acα2−6Galβ1−4Glc/GlcNAc	427

[a] Systematic names and common names of organisms are according to the GenBank Taxonomy Database (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/) and related web-based resources,and not always in complete accord with the original literature. When appropriate data is not found in the above database(s), the names were left as reported in the original literature. [b] The asterisks (*) indicate the criteria used in the original literature to claim sialic acid recognition by the lectin: *, sialidase treatment of cells (erythrocytes) reduce lectin binding (hemagglutination); **, asialo-counterpart of the glycoconjugate (used as competitive inhibitor) is less effective than intact one in inhibiting the lectin binding (hemagglutination); ***, sialic acid (or other mono/oligosaccharide) inhibits lectin binding (hemagglutination). No asterisks are attached when more than one criteria was used to prove sialic acid recognition or more detailed analysis was done on the binding specificity. Sia = sialic acid, type not specified. In cases where the specificity is indicated only as Sia, binding of the lectin to sialic acids was determined by criteria * and/or ** as above. In cases in which the evidence is even less robust, we list the item with a question mark (?). [c] Sialic acid significantly inhibits lectin binding (hemagglutination), but sialidase treatment of cells (erythrocytes) does not inhibit or even enhance binding (hemagglutination).

electrolyte imbalance. Other notable examples of sialic acid-dependent toxins are those from *Clostridium tetani* and *Clostridium botulinum*, causing tetanus and botulism, respectively.

3. Protozoa

A well-studied example is EBA-175 of *Plasmodium falciparum* (the protozoa which causes the most virulent form of malaria).[108] EBA175 is suspected to be involved in plasmodium infection of erythrocytes, but there are some laboratory and field strains which do not require sialic acid for infection.[109−111] A newly discovered *P. falciparum* protein (BAEBL) is highly similar to EBA-175 and shows similar properties including sialic acid binding.[112] There are some other sialic acid-binding lectins expressed by various protozoal pathogens, such as *Tritrichomonas foetus*[113] and *Entamoeba histolytica*.[114]

B. Vertebrate Endogenous Lectins

Needless to say, if sialic acids only serve as recognition sites for dangerous pathogens, the deuterostome lineage of animals would be better off without them. One reason that vertebrates cannot afford to abandon sialic acids is that these animals express their own (endogenous) lectins that recognize sialic acids, many of them found in the immune system.

1. Complement Factor H

This first reported example of a vertebrate sialic acid-recognizing lectin is part of the alternative pathway of complement, one of the earliest response components of the innate immune system. The protein consists of 20 short consensus repeats (approximately 60 amino acids per repeat unit, also called "sushi" domains). Some of the repeat units are involved in sialic acid recognition[115] and require an intact unsubstituted sialic acid side chain (C7−C9).[94] Factor H is a regulatory (inhibitory) factor, and binding to cell surface glycoconjugates containing sialic acids prevents the alternate complement pathway from inadvertently attacking "self" cells.[116] Foreign cells which are not covered with sialic acids will not be protected by factor H and are hence exposed to the attack by complement. Thus, pathogenic microbes expressing sialic acids have a better chance of evading immediate attack by the complement pathway and tend to be more virulent.[115]

2. Selectins

Selectins are a family of three C-type lectins (defined by a shared structural motif and calcium requirement for carbohydrate recognition) which recognize particular carbohydrate structures containing sialic acids and fucose called sialyl Lewis x (Le^x) and sialyl Lewis a (Le^a), with or without sulfation at adjacent residues.[117,118] E-selectin is expressed on activated endothelial cells, L-selectin on leucocytes, and P-selectin on activated platelets or endothelial cells. All three are involved in leucocyte trafficking,[119] e.g., initiation of leucocyte localization at the site of inflammation (E- and P-selectin) and leucocyte homing to lymph nodes (L-selectin). P-selectin has also been shown to be involved in tumor metastasis.[120]

Although the carbohydrate-dependent nature of ligand recognition by selectins is established, they appear to interact with a limited range of preferred glycoprotein ligands in vivo. For example, P-selectin selectively interacts with P-selectin glycoprotein ligand-1 (PSGL-1), which expresses not only sialyl Le^x but also a cluster of three nearby sulfated tyrosine residues which enhance the binding.[121] Crystal structures of E- and P-selectins with ligands have been resolved,[122] providing insight into the basis of such specific interactions.

Selectins also recognize other (poly)anionic carbohydrates, such as heparan sulfate,[123] although the biological significance of such "alternative ligands" is not clear. The phenotypical similarities between the mice deficient in selectins[124] and those deficient in a fucosyltransferase (FucT VII) which catalyzes a crucial step in the synthesis of sialyl Le^x and sialyl Le^a [125] suggests that the selectin interactions with nonsialylated ligands are not prominent under laboratory conditions. However, exogenously administered heparin does reduce tumor metastasis in a P-selectin-dependent mechanism.[126]

Selectin orthologs have been identified in several vertebrates. There is a putative insect (*Drosophila*) selectin ortholog called *furrowed*[127] which shares a similar overall domain structure with selectins (C-type lectin domain + multiple "sushi" repeats). However, the C-type lectin domain is not quite orthologous to those of the selectins (showing higher sequence similarity to other C-type lectin domains expressed in mammals), and there are no other definite selectin orthologs in the invertebrate genomes currently available for evaluation.

3. Siglecs

Siglecs (sialic acid-binding Ig superfamily lectins) are the largest family of mammalian sialic acid-recognizing lectins.[128,129] Siglecs share a common domain structure with an amino-terminal V-set Ig-like domain, variable numbers (1−16) of C2-set Ig-like domains, a single-pass transmembrane domain, and a cytosolic tail (V-set/C2-set means that these domains are similar to variable (V) and constant (C) domains of immunoglobulin molecules). All Siglecs other than Siglec-1 are known to have cytosolic phosphorylation sites involved in intracellular signal transduction.[129]

In humans, 11 functional members (Siglecs-1−11) and one Siglec-like molecule (Siglec-L1) have been cloned and shown to be expressed on specific cell types, such as macrophages (sialoadhesin/Siglec-1), B-cells (CD22/Siglec-2), myeloid precursors/monocytes (CD33/Siglec-3), and oligodendroglia and Schwann cells (myelin-associated glycoprotein/Siglec-4a). Many Siglecs (Siglecs-1, -3, -5, -7−10) are expressed on the cells involved in innate immunity, such as monocytes, macrophages, NK cells, and granulocytes.[129] The number of Siglec genes appears to be different between species, e.g., mice have only eight Siglec genes.[130]

Biological functions of some of the Siglecs have been elucidated through gene disruption: Siglec-2/CD22 is involved in the regulation of B-cell activation[131−134] and Siglec-4a/MAG in the maintenance of myelin sheath structure.[135] X-ray crystallography of the first Ig-like domain of Siglec-1 in complex with its ligand (Neu5Acα2−3Galβ1−4Glc) showed that the side chains of three amino acid residues highly conserved among Siglecs are in direct contact with sialic acid.[136] This confirmed the importance of these residues in ligand recognition predicted by the preceding site-directed mutagenesis studies.[137−139]

There are more Siglec pseudogenes than there are functional Siglec genes in human genome,[140] suggesting that numerous gene duplications and subsequent loss of function due to inactivating mutations have taken place. This fact may have some relevance to the ongoing evolutionary arms race between vertebrate hosts and pathogens, as discussed in section III.C. A putative Siglec-4a/MAG ortholog Schwann cell membrane protein (SMP or SCMP, Siglec-4b) in birds (quail and chicken) is reported.[141] There are some database entries of cDNAs from the African clawed frog *Xenopus laevis* which show significant similarity to (and share conserved amino acid residues with) known mammalian Siglecs. Thus, at least some members of the Siglec family may exist universally in vertebrates. There are no identifiable Siglec orthologs in the invertebrate genomes currently available for evaluation. Thus, it is possible that the sialic acid-binding features of Siglecs evolved after the emergence of prominent expression of sialic acids in deuterostomes.[142]

4. Other Endogenous Lectins That Can Recognize Sialic Acids

There are some other endogenous vertebrate molecules reported to recognize sialic acids, such as CD83[143] and a neural adhesion molecule L1.[144] The former belongs to immunoglobulin superfamily but does not have the defining amino acid residues of Siglecs in its primary sequence. The latter (and many L1-related family molecules) contain a very short amino acid sequence motif YxFR (similar to the sialic acid binding site in many Siglecs) in one of fibronectin type III domains proposed as the putative sialic acid-binding site.[144] Some of the interleukins (interleukins-1α, -1β, -4, and -7) are also shown to exhibit specific sialic acid recognition.[32] However, these cytokines have no sequence homology to other known sialic acid-binding lectins. Other examples of possible sialic acid-binding lectins in vertebrates include laminin,[145] calreticulin,[146] sarcolectin,[147] and calcyclin,[148] although some of these were found to have rather broad specificity toward acidic sugars (e.g., hexose phosphates).[147]

C. Why Do Sialic Acids Have so Many Structural Variations?

There are at least two major ways to explain the evolution of structural variations in the sialic acids: endogenous and exogenous recognition processes (Figure 3). Probably both of these are valid to some degree and together represent the major driving force behind the generation of this remarkable chemical diversity. Each of these processes are considered below.

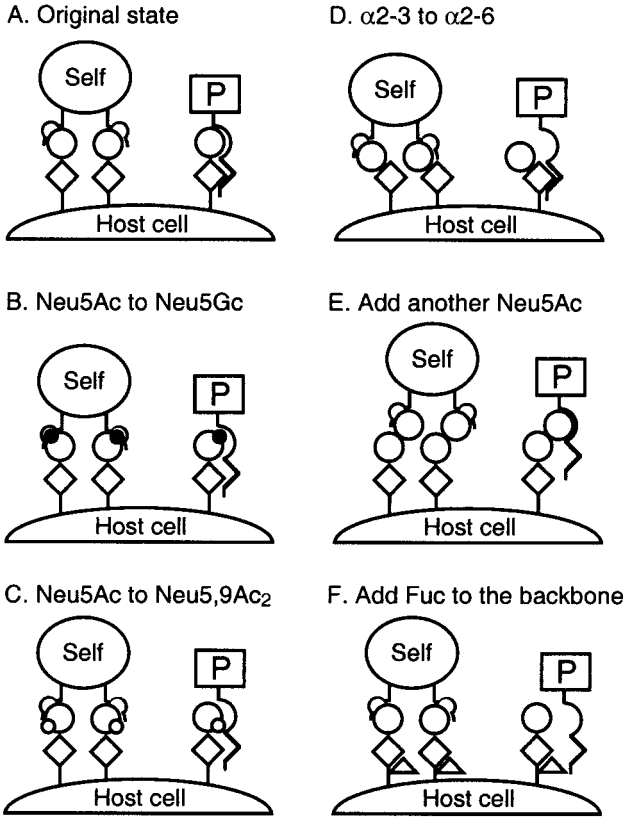


Figure 3. Modifications of sialic acids and effects on the recognition by endogenous/exogenous lectins. (A) Default—both endogenous (left) and exogenous (right) lectins can recognize sialic acids. If the pathogen lectin recognizes only sialic acids, (B) substitution at the C5 position (e.g., from Neu5Ac to Neu5Gc) or (C) substitution at the hydroxyl group(s) (e.g., from Neu5Ac to Neu5,9Ac$_2$) may prevent recognition by the pathogen lectin. If the pathogen lectin shows extended specificity toward proximal sugars, (D) changing linkage of sialic acid (e.g., from α2−3 to α2−6), (E) adding another sialic acid residue (e.g., from GM3 to GD3), or (G) adding another sugar at the proximal glycan (e.g., from sialyllactosamine to sialyl Lewis^x) may be also effective at blocking recognition.

1. Specific Recognition by Endogenous Lectins—A Driving Force from Within

One straightforward explanation for the diversity and complexity of sialic acids would be that endogenous sialic acid-binding lectins have different specificities toward different sialic acid structure ("lock-and-key"), and the preferred structure must be expressed at the proper target site. For example, mouse Siglec-2/CD22 prefers Neu5Gc to Neu5Ac,[138] while mouse and human Siglec-1/sialoadhesins prefer Neu5Ac.[149] Even more specific interaction was observed between interleukin-4 and Neu5Ac 1,7-lactone.[32] Also, modification of sialic acid may "mask" the sialic acid residue which would be otherwise recognized by endogenous lectins, as proposed for 9-*O*-acetylation which blocks recognition by Siglec-2/CD22[93] and lactamization of Neu (Neu1,5-lactam) which prevents recognition by L-selectin.[31] There may well be other examples yet to be discovered of such

specific positive and negative regulation of endogenous lectin recognition by sialic acid modifications.

The complexity of sialic acids in echinoderms could also be explained by the need for specific recognition during fertilization.[150] These animals release gametes (sperms and eggs) into open-water environments where they need to interact in a species-specific manner to avoid creating infertile hybrids. One way to prevent cross-species hybridization is synchronized release of gametes within a defined time window which is different from that of other species ("time-sharing"). However, this would work even more efficiently if these sperm and eggs have a species-specific receptor−ligand pair. On the other hand, sialic acid modifications could possibly drive speciation of echinoderms: gametes from an individual that acquired a new sialic acid modification may be able to interact with only a fraction of gametes from the same species with relaxed recognition specificity and offspring of such mating may over generations develop into a subspecies and eventually a new species. In this scenario, sialic acid−gamete lectin (in)-compatibility would work as a reproductive barrier. In accordance with these hypotheses, gametes of different species of sea urchins reportedly express glycosphingolipids unique to each species.[151]

2. Evasion of Pathogens with Sialic Acid-Recognizing Lectins—A Driving Force from Outside

Numerous pathogens recognize host sialic acids (see section III.A and Table 2). Perhaps because of their high rate of replication (allowing rapid evolution) and the survival and propagation of only the most efficient and successful infectious units, pathogens tend to evolve lectins with relatively strict specificity for sialic acid types and linkages. In contrast, we have suggested that endogenous lectins need to tolerate a certain degree of structural fluctuation of ligands.[152] This would allow evasion of pathogens by changing modifications and linkages of sialic acids on surfaces exposed to the pathogens, all the while maintaining normal internal functions (Figure 3). Even a decreased rate of infection (partial immunity) may benefit the host by reducing the severity of disease and limiting the spread of the pathogen through the population. Also, pathogens, by virtue of their shorter generation time, can diversify much faster than more complex hosts with longer generation times. As many pathogens evolve different binding specificities, the hosts would have to counter with various further modifications, thus driving structural diversification of the sialic acids. This notion conforms to an evolutionary concept called the "Red Queen Effect",[153] after the Red Queen's words to Alice in *Through the Looking Glass*, "it takes all the running you can do, to stay in the same place". Of course, averting one pathogen by a particular type of sialic acid modification could actually make the host more vulnerable to another (for example, 9-*O*-acetylation of Neu5Ac prevents recognition by influenza A virus, but this very structure is the preferred ligand for influenza C virus). Since evolution does not occur in a goal-oriented way, host organisms are probably forced to diversify themselves by a shotgun approach involving a certain amount of "waste", which might make up some part of the diversity in sialic acids in a given cell type of a given species.

There is no direct experimental evidence to prove this hypothesis. It is hard to trace evolutionary changes of sialic acids in a multicellular host species in relation to multiple pathogens with well-defined sialic acid binding specificity. Aside from the complexity and population size needed for such an experiment, we simply do not live long enough to witness the evolution of such multicellular organisms. However, some circumstantial evidence appears to support the idea. For example, the efficiency of viral infection is affected by the type of sialic acid expressed on host cell surface, e.g., the case of influenza A virus as described above. The prominent location of sialic acid 9-*O*-acetylation on mucosal surfaces and apparent changes in response to infection[154,155] suggest that sialic acid acetylation may have originated as a countermeasure against pathogens. In support of this notion is also the fact that such modified sialic acids tend to be relatively or absolutely resistant to various microbial sialidases.[156−158] On the other hand, many endogenous sialic acid-binding lectins are associated with the immune system and appear to be relatively tolerant of sialic acid structural variations.[129,159] These observations appear to support the hypothesis discussed above—that endogenous receptors with relaxed recognition specificity are more permissive to sialic acid modification, making it possible to evade pathogens by modifying sialic acids without compromising endogenous function. Of course, if the binding specificity of a particular endogenous lectin is too strict and not permissive to modifications, such a lectin gene might be inactivated and/or replaced by a new lectin with more relaxed specificity. Some of the numerous Siglec pseudogenes in the human genome could represent the results of such selection events.

Although the general direction of evolution of sialic acids (or for that matter of any evolutionary process) tends toward diversification, the same force may sometimes cause loss of variety if the situation allows it. For example, if a host can get rid of a particular type of sialic acid (SiaX) and thereby reduce susceptibility to certain types of pathogen(s) (without markedly affecting endogenous lectin recognition), then individuals with the genetic mutation causing such loss may have better fitness (survival/reproductive success) and the mutation will be positively selected and eventually fixed in the whole species. In addition, since the hosts can now generate antibodies against SiaX, they may even protect themselves from pathogens transmitted from other species carrying such SiaX. For example, the loss of Neu5Gc expression in humans[160,161] and the sporadic presence of anti-Neu5Gc antibody in human blood[162] could be related to a zoonotic Neu5Gc-binding pathogen which threatened human survival in the past.

Applying the same logic, the rather radically diverse "sialic acids" repertoire in bacteria (Neu5Ac, KDN, legionaminic acid, and pseudaminic acid) may be explained by their evolutionary race against

bacteriophages, the viruses that infect bacteria. It is known that the host range of bacteriophage is often determined by the polysaccharides expressed by the host bacteria, and phage NM8, which infects and lyses a soil bacteria *(Sino)rhizobium meliloti* M11S strain, is shown to utilize the bacterial lipopolysaccharides containing sialic acid as its attachment site.[77] Rhizobia are known to express various sialic acids (Neu5Ac, KDN, and pseudaminic acid derivatives) as well as KDO as components of their extracellular polysaccharides.[78] Hence, the diversity of sialic acids in this group of bacteria might represent an escape mechanism from bacteriophage infections.

Alternative linkages (α2−8 or α2−9) and sporadic *O*-acetylation of Neu5Ac found in polysaccharides produced by some pathogenic bacteria, such as *E. coli* K1 and *N. meningitidis* group C,[65] might be a compromise between the necessity for these bacteria to counter the bacteriophage threat while avoiding detection by the host's immune system, which should easily detect unusual sugars such as legionaminic acid and pseudaminic acid. Indeed, Endo-N, an enzyme expressed by bacteriophage phi 92 (a lytic phage of *E. coli* K92 strain) which catalyzes hydrolysis of α2−8 linked poly(Neu5Ac) on the bacterial cell surface, is unable to hydrolyze α2−9 linked poly-(Neu5Ac).[163] There are also examples of bacteriophage evasion by bacteria involving acetylation of surface polysaccharides.[164]

IV. Biosynthetic Pathways for Sialic Acids and Other α-Keto Acids

As mentioned earlier, from the biosynthetic point of view, all sialic acids (excluding certain bacteria-specific sialic acid-like sugars) are derivatives of either Neu5Ac or KDN. Although the biosynthetic pathway of KDN appears to be very similar to that of Neu5Ac,[6−8] the enzymes involved are not as well characterized; hence, the discussions below will focus more on the biosynthesis of Neu5Ac. This pathway in bacteria and vertebrates was elucidated in the 1960s, largely by efforts of the Roseman and Warren groups (reviewed by E. J. McGuire in chapter 4 of ref 165). The genetic basis for sialic acid expression in the *E. coli* K1 strain was extensively studied over the next few decades (reviewed in refs 65 and 166) and later in *N. meningitidis*.[167,168] However, the enzymes catalyzing these reactions in vertebrates were only recently cloned. It is clear that the steps of Neu5Ac biosynthesis in bacteria and vertebrates are remarkably similar (Figure 4). In addition, this pathway is very similar to that of KDO, an acidic sugar commonly found as a core component of lipopolysaccharide in Gram-negative bacteria. This biochemical similarity is reflected in the sequence similarity of some of the enzymes involved in the biosynthesis of these sugars, as discussed later.

A. Biosynthesis of Neu5Ac

As for most sugars, the ultimate source of Neu5Ac is glucose. The unique portion of pathway for Neu5Ac synthesis begins from epimerization of *N*-acetylglu-

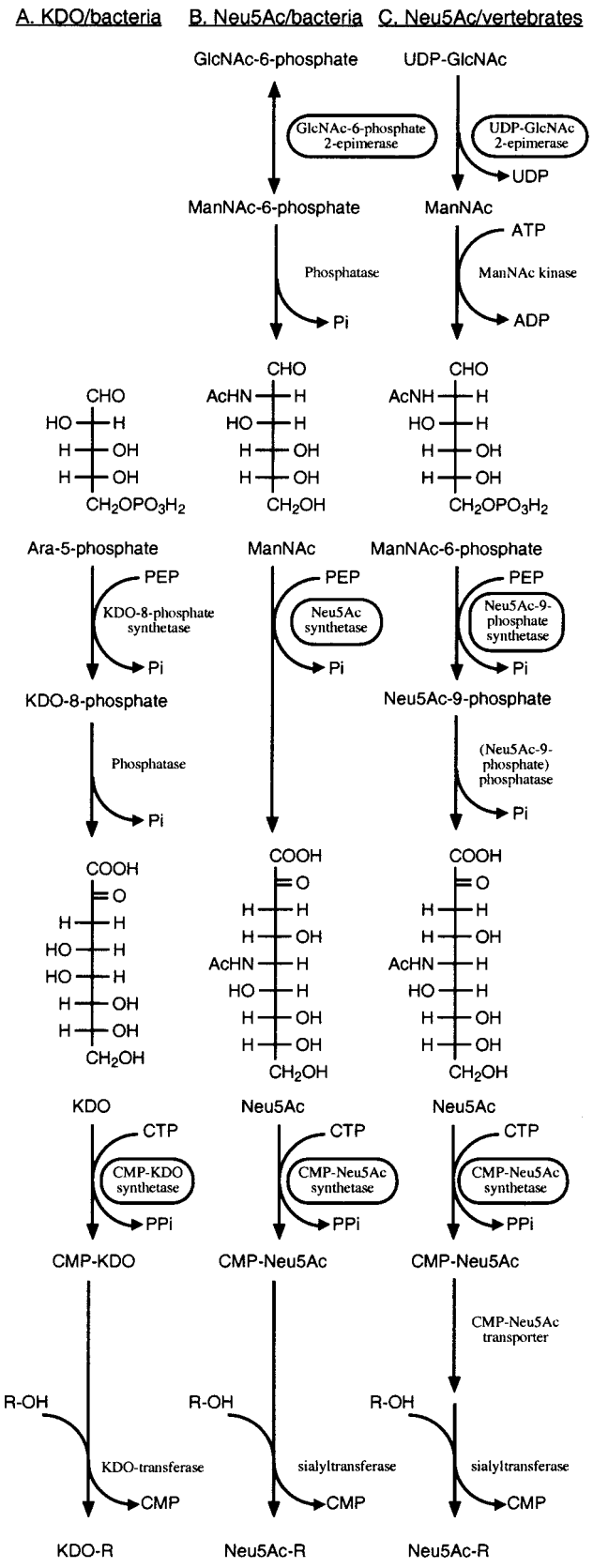


Figure 4. Pathways for Neu5Ac and KDO expression. (A) KDO expression in bacteria. (B) Neu5Ac expression in bacteria. (C) Neu5Ac expression in vertebrates. Enzymes showing significant sequence similarity are boxed. Vertebrate UDP-GlcNAc 2-epimerase/ManNAc kinase is a bifunctional enzyme with characteristics of a fusion protein: the N-terminal half is a UDP-GlcNAc 2-epimerase and C-terminal half is a ManNAc kinase.[170]

cosamine (GlcNAc) to *N*-acetylmannosamine (ManNAc). Hence, we start the following discussion with the enzymes catalyzing this reaction.

1. Synthesis of ManNAc(-6-phosphate)

In vertebrates, a bifunctional enzyme UDP-GlcNAc 2-epimerase/ManNAc kinase first converts UDP-GlcNAc to ManNAc and the ManNAc is immediately phosphorylated at the C6 hydroxyl group to yield ManNAc-6-phosphate.[169,170] Alternatively, ManNAc-6-phosphate can be synthesized by phosphorylation of ManNAc by GlcNAc/ManNAc kinase if free ManNAc is available.[171] In bacteria, it is believed that GlcNAc-6-phosphate is epimerized to ManNAc-6-phosphate by GlcNAc-6-phosphate 2-epimerase (*NeuC* in *E. coli* K1 strain, *SiaA* in *N. meningitidis*)[172,173] and then dephosphorylated by a phosphatase. However, there is a claim that ManNAc synthesis in bacteria is similar to that in vertebrates, involving UDP-GlcNAc epimerization and hydrolysis.[174] Regardless, there is clear sequence similarity between bacterial and vertebrate epimerases, suggesting that these originated from the same ancestral gene. Interestingly, many bacteria possess another GlcNAc 2-epimerase (UDP-GlcNAc 2-epimerase, which catalyzes interconversion of UDP-GlcNAc and UDP-ManNAc; *WecB* in *E. coli*)[175] which shows significant similarity to vertebrate UDP-GlcNAc 2-epimerase but not to the bacterial GlcNAc-6-phosphate epimerase. Perhaps these genes were duplicated a long time ago in bacteria and diverged to such an extent that the sequence similarity is no longer recognizable.

2. Synthesis of Neu5Ac

In vertebrates, ManNAc-6-phosphate is condensed with phosphoenolpyruvate to yield Neu5Ac-9-phosphate (the reaction catalyzed by Neu5Ac-9-phosphate synthetase[176,177]) and the phosphate group is then removed by a phosphatase which is yet to be cloned. In bacteria, ManNAc and phosphoenolpyruvate are condensed to yield Neu5Ac, the reaction being catalyzed by a Neu5Ac synthetase (*NeuB* in *E. coli* K1 strain, *SiaC* in *N. meningitidis*).[178] The similarity between bacterial and vertebrate enzymes is high—in fact, the vertebrate enzymes were cloned based on the sequence similarity to the bacterial enzyme.[176,177] Although sialic acid aldolase (*NanA* in *E. coli*) can also synthesize Neu5Ac by aldol condensation of pyruvate and ManNAc,[165] the occurrence of Neu5Ac in bacteria correlates better with the presence of Neu5Ac synthetase than with the aldolase, suggesting that the synthetase plays the major role in Neu5Ac production. Interestingly, the sialic acid aldolase shows very high sequence and structural similarity to dihydrodipicolinate synthetase (an enzyme involved in lysine biosynthesis),[179] suggesting that it is a rather novel evolutionary invention derived from the latter. Some mammalian aldolase sequences are available in GenBank (accession number AJ271330, AK002734, AF338436) and show significant similarity to the bacterial enzymes.[180]

3. Synthesis of CMP-Neu5Ac

In both bacteria and vertebrates, free Neu5Ac is converted to the active form CMP-Neu5Ac using cytidine 5′-triphosphate (CTP), a reaction catalyzed by CMP-Neu5Ac synthetase (*NeuA* in *E. coli* K1 strain, *SiaB* in *N. meningitidis*).[181,182] In vertebrates[183] the synthesis appears to occur in nuclei, and the product is then transported via a CMP-sialic acid transporter[184] into the Golgi apparatus, where glycoconjugates are assembled. Bacteria do not require such a transporter, since they lack intracellular organelles. The bacterial and vertebrate CMP-Neu5Ac synthetase are again remarkably similar,[183] although most of the bacterial orthologs appear to be shorter and show similarity to only the N-terminal half (~220 amino acids) of the vertebrate enzyme. In other words, the vertebrate enzyme appears to be a fusion protein of CMP-sialic acid synthetase to another protein of unknown function.

4. Transfer of Sialic Acids to Glycoconjugates by Sialyltransferases

The donor CMP-Neu5Ac is used by sialyltransferases to add Neu5Ac to acceptor substrates. In mammals, 18 different types of sialyltransferases have been cloned so far: one for the Siaα2−6Gal linkage, six for Siaα2−3Gal, five for Siaα2−8Sia, and six for Siaα2−6GalNAc (reviewed in refs 185 and 186; see ref 19 for nomenclature). These enzymes all share certain sequence motifs named "sialylmotifs",[187] suggesting their common ancestry. While showing strict acceptor substrate specificity, their donor substrate specificity can be somewhat relaxed, e.g., many sialyltransferases can utilize CMP-Neu5Gc and some can utilize CMP-Neu5,9Ac_2 as well[188] (the biological significance of the latter fact is unclear, since *O*-acetylation of sialic acids takes place after incorporation of sialic acids into glycoconjugates). At least one sialyltransferase (ST6Gal I) was shown to also utilize CMP-KDN.[189] On the other hand, there appears to be a specific KDN-transferase in rainbow trout.[7] Information about the rest of the transferases needed to synthesize the known KDN linkages is awaited.

In contrast to the apparent common ancestry of the mammalian sialyltransferases, at least five unrelated classes of sialyltransferases have been reported in bacteria: α2−8(9) polysialyltransferases of *E. coli* K1/K92 strains and *N. meningitidis* groups B/C,[190] α2−6 sialyltransferase of a marine bacterium *Photobacterium damsela*,[191] α2−3/6 sialyltransferase of *N. meningitidis* and *N. gonorrhoeae* involved in lipooligosaccharide synthesis,[59,192] α2−3 sialyltransferase of *Haemophilus ducreyi*/*H. influenzae*/*Streptococcus agalactiae*,[57,193] and α2−3 sialyltransferase of *C. jejuni*.[58] Interestingly, none of these bacterial sialyltransferases shows significant similarity to vertebrate enzymes either, suggesting that bacterial and vertebrate enzymes were invented independently. The lack of sequence similarity even among the bacterial enzymes suggests that sialyltransferases were reinvented many times in bacteria.

B. Can Genomic DNA Sequences Tell Us about the Presence of Sialic Acid?

With the nucleotide sequence of the enzyme genes in one hand and the whole-genome sequences in the

other, we can now rapidly look for putative biosynthetic mechanisms for Neu5Ac in a vast array of organisms. Such a survey should be even more sensitive than a screening approach employing biochemical analysis. Even if the expression of a given gene is difficult to detect because it is under strict spatio-temporal regulation, the genomic DNA will always contain the gene. In addition, immediate access to the sequence information enables us to analyze the phylogenetic relationship of these genes, as discussed below. Of course such predictions should be considered as a mere screening process, and the actual presence (or absence) of sialic acid and enzymes involved in sialic acid biosynthesis must be eventually verified by biochemical and/or genetic analysis. Meanwhile, it is intriguing that such screening already gives us some insight into the evolution of enzymes involved in sialic acid biosynthesis. Since ManNAc is a rather common sugar in bacteria (as in Gram-positive bacteria or as a precursor of *N*-acetyl-D-mannosaminuronic acid, the enterobacterial common antigen in Gram-negative bacteria[194]), the presence of Neu5Ac(-9-phosphate) synthetase and CMP-Neu5Ac synthetase genes seems to be good criteria to predict the presence of sialic acid. There are of course exceptions: *H. influenzae*, which does not have Neu5Ac synthetase but has a CMP-Neu5Ac synthetase, utilizes free Neu5Ac from host humoral environment (transported via sialic acid permease) to sialylate its lipooligosaccharide;[195] *N. gonorrhoeae*, which lacks both enzymes, still sialylates its lipooligosaccharide via a sialyltransferase that uses the minute amount of free CMP-Neu5Ac in the host humoral environment.[196] Thus, in scanning genome sequences, the absence of evidence should not be considered as evidence of absence.

1. Methanococcus jannaschii—An Archaeal Organism with Sialic Acid?

The genome of a hyperthermophilic archaea *M. jannaschii*[197] contains DNA segments similar to Neu5Ac(-9-phosphate) synthetase and CMP-Neu5Ac synthetases (these putative genes are numbered as MJ1065 and MJ1063, respectively). The presence of these genes in a cluster presumably involved in oligo/polysaccharide synthesis suggests that they are expressed together and synthesize glycan chains containing sialic acid or a similar molecule. However, there has been no direct report on the presence of sialic acid in this or any other archaeal organism. Such findings could imply that the origin of the sialic acid biosynthesis pathway is quite old, even predating the split of bacteria and archaea/eukarya clades (i.e., the common ancestor of cellular life might have been capable of expressing sialic acids). On the other hand, there are no similar genes in any of the other archaeal genomes currently available. Thus, it could also be that *M. jannaschii* acquired these genes by lateral transfer.

2. Legionella pneumophila—Sialic or Not Sialic?

L. pneumophila (a bacterium causing Legionnaire's disease, a serious pulmonary infection in immunocompromised individuals) expresses a homopolymer of legionaminic acid (Figure 1C) as a distal part of its lipopolysaccharide (LPS).[198] Recent studies have identified the segment of *Legionella* genome involved in the biosynthesis of LPS,[12] which contains putative genes highly similar to the three genes involved in the synthesis of Neu5Ac, namely, GlcNAc-6-phosphate 2-epimerase, Neu5Ac synthetase, and CMP-Neu5Ac synthetase. While this bacterium does not express Neu5Ac, these putative genes were shown to be able to complement *E. coli* K1 mutants deficient in these enzymes. From these results the authors concluded that the biosynthesis of legionaminic acid is similar to that of Neu5Ac.[12] Although it is possible that the precursor of legionaminic acid is a sugar other than ManNAc, e.g., derivatives of quinovosamine (2-amino-2,6-dideoxy-D-glucose) or bacillosamine (2,4-diamino-2,4,6-trideoxy-D-glucose), the origin of enzymes involved in the biosynthesis of Neu5Ac and legionaminic acid appears to be common.

3. Drosophila melanogaster—Incomplete Tool Kit?

The fruit fly *D. melanogaster* was one of the first multicellular organisms to have its genomic DNA completely sequenced.[199] Although our initial assessment of the *Drosophila* genome did not show any positive evidence for the Neu5Ac biosynthetic machinery,[142] a more thorough analysis provides us a somewhat different picture. Although the *Drosophila* genome indeed lacks orthologous genes of enzymes for synthesis of ManNAc (such as GlcNAc-6-phosphate 2-epimerase, GlcNAc 2-epimerase,[200] UDP-GlcNAc 2-epimerase/ManNAc kinase), it appears to contain several other putative genes possibly involved in the downstream aspects of the pathway. For example, DNA segments similar to the genes of such proteins as GlcNAc/ManNAc kinase[171] (GenBank accession number of the putative protein: AAF55174), Neu5Ac(-9-phosphate) synthetase (AAF54811), CMP-Neu5Ac synthetase (not annotated), CMP-sialic acid transporter (AAF45808, possibly another sugar nucleotide transporter), and one sialyltransferase (ST6Gal I-like, AAF47256) are found in the *Drosophila* genome. Thus, it is possible that if ManNAc is provided by an external source and the expression of all the other enzymes coincides at the right time in the right place, *Drosophila* may be able to express sialic acids, as previously reported.[38] The question is the origin of ManNAc: it could be from the diet, or from symbiotic/parasitic bacteria (e.g., *Wolbachia*), or due to the existence of as yet unidentified enzymes which can catalyze epimerization of GlcNAc.

4. Caenorhabditis elegans—No Sign of Sialic Acid

Another multicellular invertebrate animal that has its entire genomic DNA sequenced is the nematode *C. elegans*.[201] Earlier literature reported the absence of sialic acid in *C. elegans*.[202] In accordance with this report, there seem to be no orthologs of the genes involved in the biosynthesis of sialic acids (the enzymes of ManNAc synthesis, Neu5Ac synthetase, or CMP-Neu5Ac synthetase) in the genome of *C. elegans*.

C. Biosynthesis of KDO in Bacteria and Comparison to that of Neu5Ac

KDO is expressed in Gram-negative bacteria (as an essential component of lipopolysaccharides) and in higher plants and green algae (as a component of cell wall polysaccharides). The biosynthesis of KDO (Figure 4) begins with condensation of D-arabinose-5-phosphate and phosphoenolpyruvate to yield KDO-8-phosphate (catalyzed by *KdsA* in *E. coli*). It is then dephosphorylated and conjugated with CTP to yield CMP-KDO (catalyzed by *KdsB* in *E. coli*), providing the donor for KDO incorporation into oligo/polysaccharides by KDO-transferase (*KdtA* in *E. coli*). The similarities in chemical structures of KDO and Neu5Ac and in the reactions involved in the biosynthesis of these sugars suggests a common evolutionary origin. Indeed, the CMP-KDO and CMP-Neu5Ac synthetases show significant sequence similarity to each other. However, the rest of the genes in the pathways do not. Given the dissimilarity among bacterial sialyltransferases themselves, the dissimilarity between KDO transferases and sialyltransferases is not too surprising. However, it is rather puzzling that the KDO-8-phosphate and Neu5Ac(-9-phosphate) synthetases do not show recognizable sequence similarities. This could be a reflection of different facial selectivity of these enzyme reactions: in the KDO-8-phosphate synthesis reaction, phosphoenolpyruvate approaches from the *re* face of the aldehyde group of D-arabinose-5-phosphate (hence the configuration at C4 of the product is *R*),[203] while in Neu5Ac synthesis it is probably from the *si* face of the aldehyde group of ManNAc(-6-P) (hence the configuration at C4 of the product is *S*). Thus, we may need the three-dimensional structure of Neu5Ac synthetase and knowledge of key residues involved in the substrate recognition and catalytic reaction before recognizing similarities of the two enzymes. Meanwhile, there is in fact, such a precedent: another 2-keto-3-deoxy acid, 3-deoxy-D-*arabino*-heptulosonic acid 7-phosphate (DAHP, an intermediate in shikimic acid pathway, which produces aromatic amino acids in bacteria, plants, and fungi), is synthesized from D-erythrose-4-phosphate and phosphoenolpyruvate by DAHP synthetase via a very similar reaction to KDO-8-phosphate synthesis. However, the structural and sequence similarities between these enzymes in *E. coli* were revealed only after the crystal structures of DAHP synthetase[204] and KDO-8-phosphate synthetase[205] were determined and conservation of several key residues at structurally orthologous positions were found.[205] Overall, we can conclude that at least some aspects of the KDO and sialic acid biosynthesis pathways are evolutionarily related.

D. Phylogenetic Relationships of Enzymes Involved in Sialic Acid Biosynthesis

In an attempt to obtain better insights into the evolutionary history of sialic acids, we constructed phylogenetic trees for two essential enzymes in sialic acid biosynthesis: CMP-Sia synthetase (Figure 5A) and Neu5Ac(-9-phosphate) synthetase (Figure 5B). These trees were reconstructed using the neighbor-joining method,[206] which tries to find the evolutionary relationships of the genes based on the sequence similarities (distances) between all possible pairs of the genes in the set. Essentially the same tree topology was obtained using another method called maximum parsimony (which tries to emulate the path of evolutionary changes while minimizing the number of changes necessary to explain the current status of the genes). As discussed below, these trees actually suggest that the genes involved in biosynthesis of sialic acids may be ancient in origin and not unique inventions in deuterostomes which were later introduced via lateral transfer to pathogenic bacteria. In fact, the genes of pathogenic bacteria do not show more significant sequence similarity to vertebrate genes than to the orthologous genes from other sources such as *Drosophila,* i.e., the branch of pathogenic bacterial genes is not rooted between the branches for *Drosophila* and vertebrates, which should be the case if lateral gene transfer(s) had occurred (see Figure 5).

1. CMP-KDO and CMP-Neu5Ac Synthetases

The significant sequence similarity of CMP-KDO and CMP-Neu5Ac synthetases allows us to analyze the phylogenetic relationship of these enzymes and their relatives in the available genome databases. As shown in Figure 5A, there are three major branches: one for CMP-KDO synthetases, one for CMP-Neu5Ac synthetases, and the third for related genes of unknown function. Because three of the five genes in this third group (*M. jannaschii*, *B. subtilis*, and *C. crescentus* 1) are physically adjacent to the Neu5Ac synthetase-like genes in the respective genomic DNA (i.e., in a gene cluster), they are likely to catalyze sugar nucleotide synthesis, with a substrate similar to sialic acid. Thus, this group probably represents part of the extended CMP-Neu5Ac synthetase family.

The CMP-KDO synthetases and CMP-Neu5Ac synthetases form mutually exclusive clades, implying that these two groups of enzymes originated from an ancestral enzyme gene by a duplication, subsequently acquired distinct functions, and then became noninterchangeable. The timing of such a gene duplication is difficult to estimate from this tree (we have too few samples from archaea and eukarya to infer it with confidence). From the fact that all three domains of life (archaea, bacteria, and eukarya) are represented in the CMP-Neu5Ac + "CMP-Sia" synthetase clade, it is tempting to propose that the duplication happened in the common ancestor of all cellular organisms. However, we cannot rule out the possibility of lateral gene transfer between bacteria and archaea[207] or bacteria and eukaryotes.[208] It is also possible that some of the genes included in the analysis are inappropriate, i.e.,"the third branch" may be irrelevant after all and the gene in cyanobacterium *Synechocystis* may not be a CMP-Neu5Ac synthetase, since this organism does not have Neu5Ac synthetase-like gene.

Peripheral branch topologies (the way branches are clustered/split) of the CMP-Neu5Ac synthetase clade provide some limited insight into the origin of sialic acids. CMP-Neu5Ac synthetase genes in animals

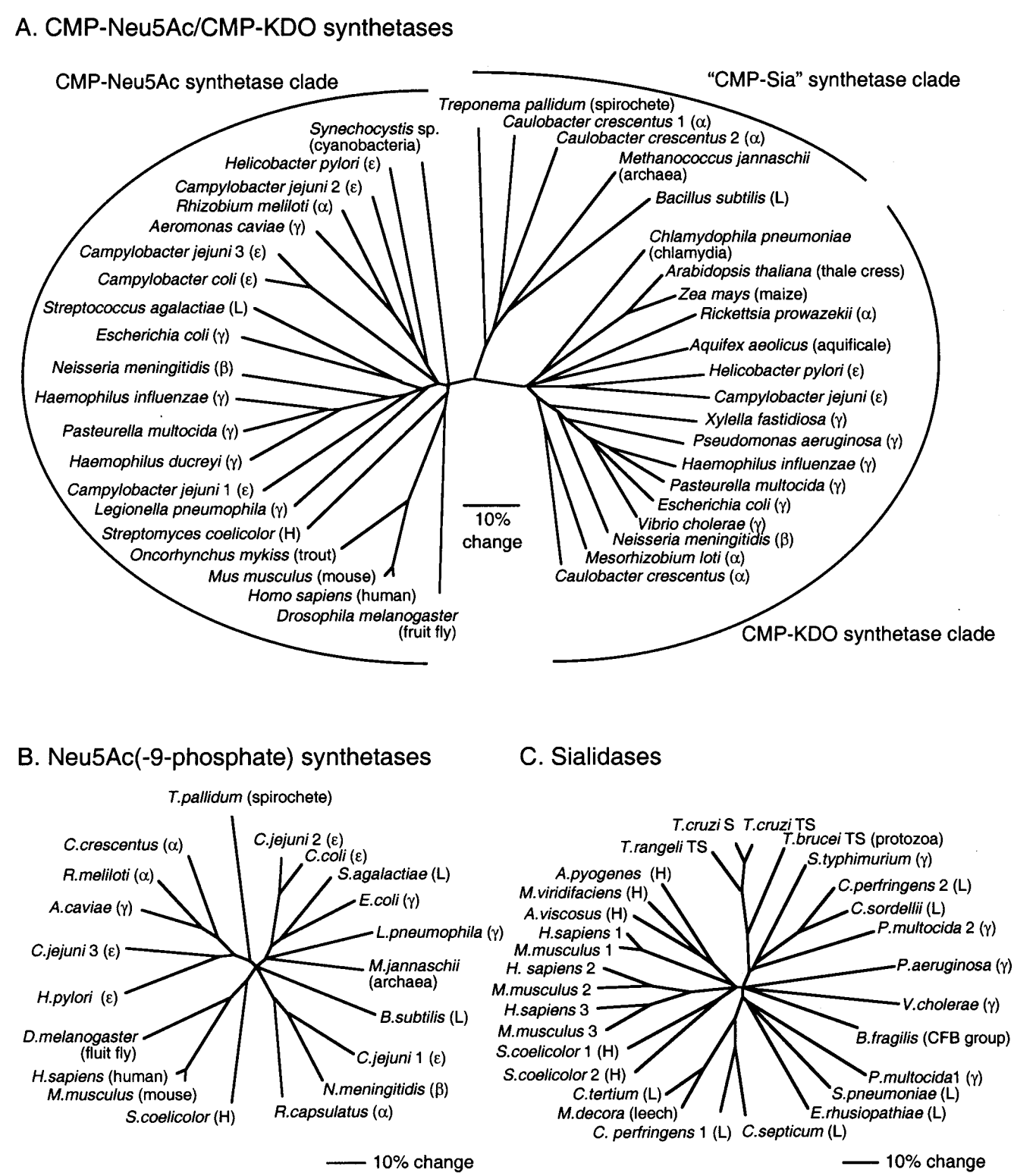


Figure 5. Phylogenetic relationships of enzymes involved in the metabolism of sialic acids. (A) CMP-Neu5Ac synthetases and CMP-KDO synthetases, (B) Neu5Ac(-9-phosphate) synthetases, and (C) sialidases from various sources. In A, the CMP-KDO group serves as the root of CMP-Neu5Ac synthetase group and vice versa. Notice that the genes from metazoan animals (fly and vertebrates) are monophyletic (form an exclusive clade) in both A and B. The association of *S. coelicolor* with metazoan animals' clade may be an indication of horizontal gene transfer (see text). The letters in parentheses following the systematic names of bacteria specify their classifications: α−ϵ, proteobacteria (Gram-negative) α−ϵ subdivisions; H, high G+C Gram-positive bacteria; L, low G+C Gram-positive bacteria. In C, trypanosoma sialidase and *trans*-sialdiases are indicated with capital S and TS, respectively. Vertebrate sialidases NEU1, NEU2, and NEU3 are indicated with numbers.

(fruit fly, rainbow trout, mouse, and human) are monophyletic (forming an exclusive branch), suggesting that the ancestor of all metazoan animals carried this gene (and probably sialic acids). Surprisingly, one of the bacteria (*Streptomyces coelicolor*, a Gram-positive bacterium) seems to have shared an ancestral gene with these animals, suggesting a possibility of lateral gene transfer between the ancestors of this

bacterium and eukaryotic animals. We cannot infer the direction of this gene transfer, because the samples from these groups of organisms are so limited and the root is so deep. Additional sequence information from sialic acid-positive fungi (e.g., *C. albicans*, *C. neoformans*, and *A. fumigatus*) and protozoa (e.g., *E. histolytica* and *D. discoideum*) may help in improving the resolution of the analysis. Whole-genome DNA sequencing of most of these organisms is currently under way.

The topology of CMP-Neu5Ac synthetase tree strongly implies lateral transfers of this gene among bacteria. For example, quite distant species of bacteria carry very similar copies of this gene (e.g., Gram-negative bacteria *E. coli* and Gram-positive bacteria *S. agalactiae*), and the different copies of this gene in a single organism shows remarkable sequence divergence, each showing higher similarity to the gene from other bacteria (*C. jejuni* strain NCTC11168). In contrast, the branching pattern for the CMP-KDO synthetase tree more faithfully reflects the phylogenetic relationships of these organisms deduced from ribosomal RNA[209] and from some other essential genes. Only a few irregularities are observed, e.g., monophyly of α-proteobacteria *Rickettsia prowazekii* and hyperthermophilic bacteria *Aquifex aeolicus* and monophyly of higher plants and chlamydia. The latter is proposed to be the evidence of bacteria-to-(proto)plant lateral transfer of the CMP-KDO synthetase gene.[210,211]

2. Neu5Ac(-9-phosphate) Synthetases

The picture provided by the phylogenetic analysis of Neu5Ac(-9-phosphate) synthetase (Figure 5B) is less informative, partly because we cannot define the root of the tree (to define the root we need an outgroup, i.e., a branch consisting of another gene which shares an ancestor with the gene of interest). However, assuming that the root is somewhere near the center of the wheel, the topologies of "peripheral" branches suggest some interesting possibilities. For example, the monophyly observed for the genes from metazoan animals (fruit fly, mouse, and human) again suggests the common ancestor of these animals had a prototype of this gene (hence sialic acids). As with CMP-Neu5Ac synthetase, the putative Neu5Ac-(-9-phosphate) synthetase of the bacterium *S. coelicolor* forms a monophyletic branch with metazoan animals, strengthening the impression that lateral transfer between the ancestor of this bacterium and an early eukaryotic organism might have happened. The bacterial branches again show irregular clustering patterns not conforming to their organismal phylogeny, suggesting that multiple lateral transfers of this gene have happened between bacteria belonging to different groups. An interesting new paper shed light on the significance of divergent multicopy Neu5Ac synthetases in *C. jejuni*, which suggests that each copy has a distinct function.[212]

V. Biosynthesis of Modified Sialic Acids

As can be seen from Table 1, most sialic acid modifications reported to date have been found in deuterostomes and a few in bacteria. Unfortunately, as of today, there is not much sequence information on the enzymes involved in generating these modifications. CMP-Neu5Ac hydroxylase,[9] which adds an oxygen atom to the acyl group of Neu5Ac to generate CMP-Neu5Gc, is the only gene in this category ever cloned. Much effort has been directed toward purifying and cloning the enzyme(s) directly involved in the transfer of acetyl groups to the hydroxyl groups of sialic acids but with limited success.[213,214] As for other modifications (e.g., 8-*O*-methylation/sulfation), the rarity of such modified sialic acids in vertebrates makes them unattractive as sources to purify or clone the enzymes involved in such modifications: echinoderms are far better suited for this purpose. However, DNA sequence information concerning echinoderms is quite limited at this time. Progress in the sea urchin genome project as well as in the sea urchin EST (cDNA) project is awaited.[215] This may give us new clues to identify enzymes involved in sialic acid metabolism and modifications and to analyze the significance of structural diversity of sialic acids.

Although the amount of information currently available for the enzymes is limited, the narrow range of organisms (echinoderms and vertebrates) expressing these modified sialic acids raises the hope that they will share a certain conserved motif involved in sialic acid recognition. If so, the cloning of one of these enzymes will lead to accelerated discovery of the rest, based on the sequence similarity.

A. CMP-Neu5Ac Hydroxylase

CMP-Neu5Ac hydroxylase (monooxygenase) is a cytosolic enzyme which introduces a single oxygen atom to the *N*-acetyl group, giving CMP-Neu5Gc.[9,216] This enzyme requires some cofactors: NAD(P)H, cytochrome b_5, and cytochrome b_5 reductase.[217,218] It is considered to be the only documented pathway to generate Neu5Gc de novo: the expression of this enzyme and that of Neu5Gc in tissues show a generally positive correlation,[9,219] and a mutation in this enzyme (a 92 base pair exon deletion universal in modern humans)[161,220] explains the almost complete lack of Neu5Gc in human tissues. It is interesting that while there is an equivalent enzyme in the starfish *Asterias rubens*,[221] there seems to be no orthologous gene in bacteria, explaining why bacteria do not express Neu5Gc. Thus, this enzyme may be a unique invention in eukaryotes. However, the active site of this enzyme contains a Rieske iron−sulfur center,[222] a motif commonly found in many prokaryotic proteins involved in redox reactions. A recent comprehensive analysis of the phylogeny of such enzymes showed that the CMP-Neu5Ac hydroxylases form a unique subgroup with a very ancient root.[223]

B. Sialic Acid *O*-Acetyltransferases

O-Acetylation of the sialic acid side chain (C7−C9 hydroxyl groups) predominantly takes place in the Golgi apparatus, and *O*-acetyltransferase enzyme(s) associated with Golgi membrane catalyze this reaction.[224,225] The fact that different types of cells show different preferences toward *O*-acetylation of specific

glycan classes and sialic acid linkages[10,224,226] suggests that there are several *O*-acetyltransferases with different acceptor substrate specificities. The other possibility is a family of adaptor proteins which modify the substrate specificity of a single enzyme. The *O*-acetyltransferase(s) use acetyl coenzyme A (acetyl CoA) as a donor, and there is uncertainity as to whether the intact donor or just the acetyl group is transferred across the Golgi membrane.[227] Regardless, these enzymes introduce acetyl groups initially to the hydroxyl groups at the C7 or C9 position of sialic acids.[224] Acetyl groups at C7 can then spontaneously migrate to the C8 and C9 hydroxyl groups under physiological conditions.[228,229] There is also evidence for an enzyme which accelerates this migration in vivo.[230] Although there are bacterial enzymes involved in 9-*O*-acetylation of the Neu5Ac side chain, the properties of the enzyme appear to be drastically different from those of vertebrate enzymes.[231]

Sialic acids containing 4-*O*-acetyl groups seem to be prominently expressed only in a limited group of animals, such as horse, donkey, guinea pig, and echidna.[232,233] A sialic acid 4-*O*-acetyl transferase activity, also associated with Golgi membrane, has been identified in guinea pig liver.[234] However, 4-*O*-acetyl groups are very difficult to detect when they are present in small amounts. This together with the identification of 4-*O*-acetyl sialic acids as receptors for a mouse hepatitis virus[95] and the antibody-based detection of putative 4-*O*-acetyl-GM3 in human colon carcinomas[235] suggests that the 4-*O*-acetyl modification is actually more widespread in vertebrates than currently presumed.

C. Sialic Acid 8-*O*-Methyltransferase

Methylation of sialic acids at the C8 position was once considered to be unique to the echinoderms. However, it appears that vertebrates too express this kind of sialic acid, although in minute amounts.[30] The enzyme-catalyzing transfer of methyl groups from *S*-adenosylmethionine to the C8 hydroxyl group of sialic acid (sialic acid 8-*O*-methyltransferase) has been identified in starfish *A. rubens*.[236] However, no genes have been cloned.

D. Other Modifications

Given the existence of such modifications as 9-*O*-lactyl and 8/9-*O*-sulfo-sialic acids, it is natural to assume the presence of sialic acid *O*-lactyltransferase and 8/9-*O*-sulfotransferase. Also, there should be an enzyme which catalyzes the transfer of an acetyl group to the glycolyl group of Neu5Gc (to make Neu5GcAc). However, nothing has been published so far on these enzymes. It is possible that lactylation of the C9 (or C7?) hydroxyl group may be catalyzed by a sialic acid *O*-acetyltransferase if lactyl-coenzyme A is available. However, the presence of such a molecule in vertebrate tissues has not been reported.

VI. Degradation of Sialic Acids

In this review we do not discuss in detail the pathways and enzymes involved in the degradation of sialic acids, e.g., sialic acid *O*-acetylesterase,[237,238] sialidase,[239−242] sialic acid transporter,[243] and sialic acid aldolase,[244] since we considered these less relevant to the evolution of sialic acid diversity. However, these enzymes and proteins are obviously of great importance to the biological functions of sialic acids. For example, deficiency in lysosomal sialidase (NEU1) and sialic acid transporter (SLC17A5) results in sialidosis[240] and sialic acid storage disorder,[243] respectively (both are neurodegenerative disorders leading to lethality).

Phylogenetic relationships of sialidases from various sources are shown in Figure 5C. Since we cannot define the root again, the picture provided by the tree is not very useful to understand the history of this class of enzyme. However, if we look at the "peripheral" branches, the vertebrate enzymes (Neu1, 2, 3) seem to form a monophyletic clade with the sialidases from actinomycetes (Gram-positive, high G+C bacteria) including, again, *S. coelicolor*. There are two other interesting anomalies: first, trypanosoma (*trans*-)sialidases are clustered together with a group of bacterial sialidases and, second, *Macrobdella decora* (leech; protostome) sialidase is clustered together with another group of bacteria (*Clostridia*). All these observations point to the possibility of lateral gene transfer. Similar interkingdom lateral gene transfers were also suggested for sialic acid aldolase between *Trichomonas vaginalis* and some bacteria[245] as well as vertebrates and some bacteria.[180]

VII. Sialic Acid Metabolism Genes Shared Only among Vertebrates and Bacteria

The International Human Genome Sequencing Consortium reported the presence of more than 200 putative human genes that appear to be shared exclusively among vertebrates and bacteria, possibly via lateral transfer.[246] A recent paper narrowed the list by applying more stringent criteria and pointed out that alternative explanations were possible even for these, including lineage-specific gene loss, sample size effects, and evolutionary rate variation.[247] Interestingly, we have noted that 7 of the 40 or so remaining candidates are involved in the sialic acid biosynthesis (CMP-Neu5Ac synthetase, UDP-GlcNAc 2-epimerase/ManNAc kinase, and GlcNAc 2-epimerase) or turnover/degradation (ganglioside sialidase, cytosolic sialidase, sialic acid-specific acetylesterase, and sialic acid aldolase, aka *N*-acetylneuraminate lyase). Thus, a pathway involving ∼0.05% of the human genome represents almost 20% of the potential examples of lateral gene transfer between vertebrates and bacteria. Although CMP-Neu5Ac synthetase should be excluded from the list (it has a putative ortholog in fruit fly genome as described above), the rest of the 7 genes appear to be primarily shared by vertebrates and bacteria and many of the latter are pathogenic bacteria (Varki, Angata, and Gagneux, unpublished observations). A related article[180] pointed out that sialic acid aldolase may be a case of lateral gene transfer, although the direction of the transfer (bacteria-to-vertebrates or vice versa) cannot be inferred. Regardless of the mechanism (lateral transfer, lineage-specific gene loss, and/or evolutionary rate variation), the very fact that these

genes are observed exclusively in vertebrates and bacteria suggests that sialic acids have been involved in an evolutionary "arms-race" between these organisms, as discussed in this review.

VIII. Concluding Remarks and Future Prospects

We have attempted here to summarize new discoveries in the field of sialic acid biology with an emphasis on the structural diversity of sialic acids, the enzymes involved in the biosynthesis of sialic acids, and the lectins that recognize sialic acids. In addition, we adopted an evolutionary perspective to try and gain insights into the biological significance of sialic acids. With regard to the diversity of sialic acid structure, we have suggested that their evolution be considered not only in the context of specific recognition by endogenous lectins, but also in light of the necessity of multicellular host organisms to constantly change to evade pathogens which utilize sialic acids on the host cell surface. We also tried to estimate how sialic acids have emerged, based on phylogenetic analyses of the enzymes involved in the biosynthesis of sialic acids. We consider our analysis inconclusive, largely due to the poor representation of eukaryotes in the phylogenetic analysis. However, some general conclusions can be stated here. First, it appears likely that at least some aspects of the KDO biosynthesis pathway of bacteria and the sialic acid biosynthesis have common evolutionary origin. Given the more conventional distribution and phylogenetic relationships of CMP-KDO synthetases in bacteria and considering the similarity of KDO-8-phosphate synthetase to DAHP synthetase involved in an essential metabolic pathway (aromatic amino acid biosynthesis), it is reasonable to speculate that KDO biosynthesis pathway was ancestral to sialic acid biosynthesis pathway. Second, the traditional explanation that sialic acids were invented in the deuterostome lineage is not supported by our analyses. Instead, our analyses suggest that the origin of the sialic acid biosynthetic pathway predates the split of protostome and deuterostome lineages, possibly even the separation of the three domains of life. On the basis of these observations, we consider that three evolutionary scenarios are compatible with the currently available data (see Figure 6). In the first scenario, the sialic acid biosynthesis pathway is evolutionarily very ancient, predating the emergence of the three major branches of life. In this scenario, the genes of this pathway were then partially or completely lost in most species in most lineages, with the exception of the deuterostomes, a few protostomes, and some prokaryotes. The apparent preservation of the pathway in microbial pathogens could then be explained by its survival advantage in vertebrate hosts and its frequent lateral transfer among bacteria (however, the bacteria which originally conserved this pathway would not have had vertebrate hosts available and may have utilized sialic acids for some other purposes, such as self-defense from bacteriophages). In the second scenario, the sialic acids were invented in bacteria after the split of the bacteria and archaea/eukarya clades and later introduced via lateral gene transfer(s) to a limited number of organisms belonging to archaea

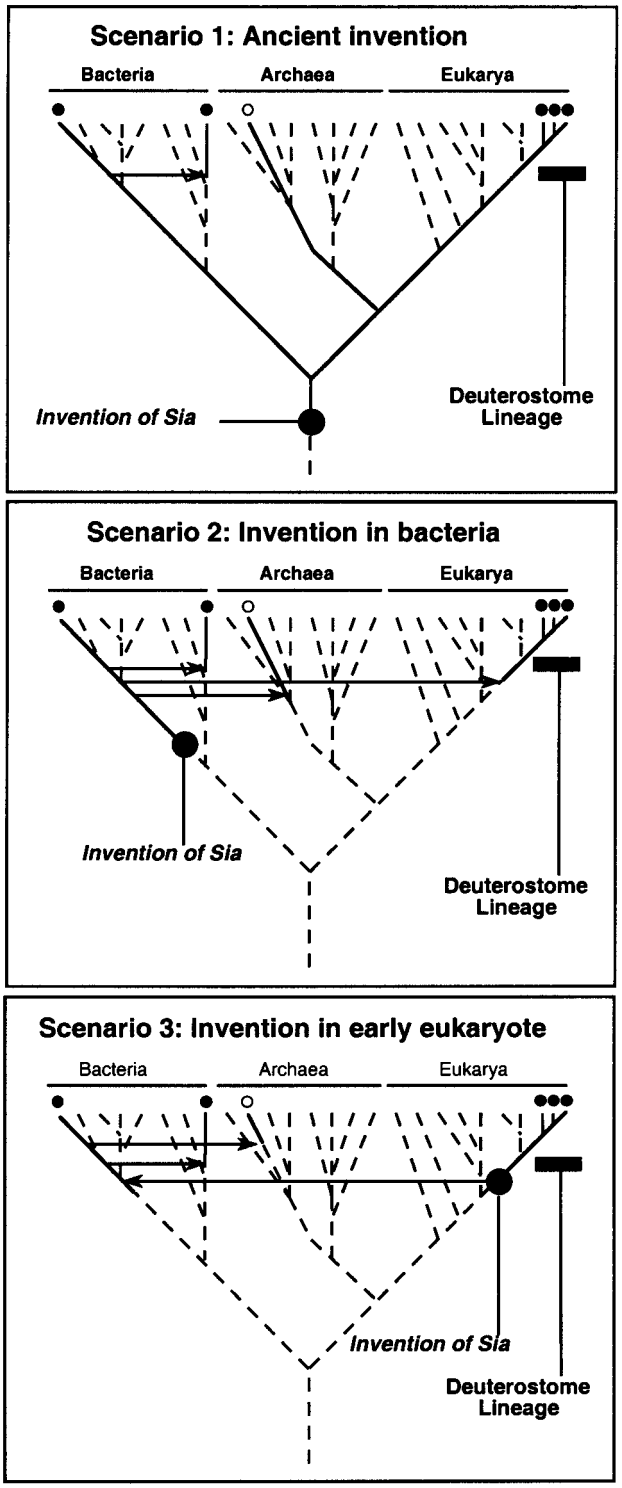


Figure 6. Possible scenarios for the invention and spread of sialic acids. Presence of sialic acids are indicated with solid circles, and possible presence of sialic acid in archaea (*M. jannaschii*) predicted from genomic DNA sequence analysis is indicated with an open circle. In scenario 1, sialic acid (or rather the biosynthetic machinery for the expression of sialic acid) was invented in the common ancestor of cellular life, inherited in all three domains of life, but subsequently lost in most of the lineages. In scenario 2, sialic acid was invented in bacteria and spread to other domains of life via horizontal (lateral) transfer. In scenario 3, sialic acid was invented in eukaryotes (prior to the split of deuterostome and protostome lineages) and spread to other domains via horizontal transfer.

and eukarya (including, perhaps, the single-cellular ancestor of eukaryotic animals). The reason we presume that a single-cellular eukaryotic organism was the recipient of bacterial DNA is that such an organism has a much better chance of incorporating foreign DNA into its genome and transmitting to its offspring, compared with a multicellular organism which has separate cell lineages (germ-line cells) set aside for reproduction.[248] Even in this scenario we have to assume extensive gene losses in many of the lineages descended from such protoanimals. In this regard, a curious fact is that a large number of protostomes seem to have soluble lectins that recognize sialic acids (see Table 4). Since many of these lectins seem to be involved in host defense, it is possible that an early ancestor in the protostome lineage found it beneficial to eliminate expression of sialic acids and instead recognize it as a foreign epitope on invading pathogens. In the third scenario, sialic acid was a unique invention in metazoan animals (deuterostomes + protostomes) and all other life forms that have sialic acids acquired the genetic tool kit from descendants of these animals. Thereafter, the pathway would have been prominently expressed in the deuterostomes and suppressed or partially eliminated in the protostomes. The occurrence of sialic acids in microbial pathogens would then have resulted either from lateral gene transfer from animals prior to the protostome:deuterostome split (as in pathogenic bacteria) or by convergent evolution involving genes that cannot be identified by homology searching (as in some pathogenic fungi). However, given the conspicuous absence of the KDO pathway in eukaryotes (except in plants which do not express sialic acids), it is rather unlikely that sialic acids were invented de novo in the immediate ancestor of metazoan animals (i.e., without any preexisting pathway serving as template).

As discussed earlier, it is interesting that a Gram-positive bacterium *S. coelicolor* (not a known vertebrate pathogen) not only has a putative Neu5Ac(-9-phosphate) synthetase and CMP-Neu5Ac synthetase, but also appears to have the enzymes involved in the degradation of sialic acids (sialidases and sialic acid aldolase, although the latter is more similar to dihydrodipicolinate synthetase). As mentioned above, *S. coelicolor* genes possibly involved in sialic acid metabolism also appear to be more closely related to those of metazoan animals than those of other bacteria. Although it is highly speculative, it could be the case that a good part of the enzymatic system of sialic acid metabolism in animals may have originated in the ancestor of this bacteria and was introduced by lateral transfer to a single-celled ancestor of metazoan animals. The most likely mechanism for such lateral transfer is via symbiosis or predation,[180,248] and this bacteria could have become a symbiont for nascent eukaryotes because of its ability to produce antibiotics.[249] The ancestor of metazoan animals lacked a complex immune system (adaptive immunity is considered to be an invention in vertebrates[250]), and the capability to produce antibiotics through symbiosis could have provided the ancestral animal (or ancestral eukaryote) with a great survival advantage. The genes involved in the sialic acid metabolism may have thus become incorporated to the eukaryotic genome as a side effect of such a symbiosis or cohabitation. Of course, the direction of such lateral gene transfer could be the reverse, and it also remains possible that the similarities in the enzymes of this bacteria and animals may be just a coincidence, due to parallel evolution or even a bias in amino acid usage. Regardless, these considerations appear to be good reasons to initiate studies of sialic acid metabolism in *Streptomyces* as well as in some lower eukaryotes, such as protozoa.

The continuing expansion of DNA sequence information in the future will make it possible to look into the deeper roots of sialic acid phylogeny with improved accuracy. Molecular cloning of many enzymes involved in the modifications of sialic acids will also be necessary to further our understanding of the mechanism and significance of structural diversity of sialic acids. Additional mouse "knock-outs" of enzymes involved in the biosynthesis and modification of sialic acids, as well as their endogenous receptors such as Siglecs, will provide more clues to understanding the importance of endogenous recognition involving sialic acids. We may also have to consider experimental systems to examine the hypothesis on the Red Queen Effect and the evolution of sialic acid diversity, perhaps employing a bacteria−bacteriophage system. These new approaches, in combination with classic chemical and biochemical approaches, will provide us new dimensions and better appreciation of the significance of the sialic acids in biological phenomena.

IX. Glossary

Gram-Positive and Gram-Negative Bacteria. Traditionally, bacteria were roughly classified by a method called Gram staining. Gram-positive bacteria generally have thicker cell walls, while Gram-negative bacteria (which are not stained) have thinner cell walls coated with a membrane of lipid bilayer. Gram-positive bacteria are further classified into two major groups: those which have high guanine (G) and cytosine (C) content (as high as 75%!) are called high G+C group (or actinobacteria), and those which have low G+C content are called low G+C group (including *Bacillus* and *Clostridium*).

Deuterostomes and Protostomes. Animals (multicellular eukaryotes with the ability to move on their own "will"; anima = soul) are broadly classified into deuterostomes and protostomes, based on the difference in their patterns of embryogenesis (embryonic development). In general, the digestive tract of protostomes is formed from the part which later becomes the mouth (proto = first, stome = mouth). In contrast, in deuterostomes the digestive tract is formed from the part which later becomes the anus, and the mouth is formed secondarily (deutero = second). The major branches of protostomes include arthropods (Tracheata = mainly insects; Chelicerata = ticks, spiders, horseshoe crabs; Crustacea = shrimps and crabs) and molluscs (Gastropoda = slugs, snails and conchs; Bivalvia = shellfishes; Cephalopoda = octopuses and squids). The major branches of deutero-

stomes include vertebrates (mammals, birds, reptiles, amphibians, fishes), ascidians (sea squirts), and echinoderms (starfish, sea cucumbers, sea urchins). The phylogenetic position of nematodes (round worms) is not clear, and these are included in protostomes in some classifications.

Genome. The original definition of genome is "a minimal set of genetic materials necessary and sufficient for the proper functioning of the organism". For cellular organisms, the physical entity which stores genetic information is DNA (hence called genomic DNA). However, it is now clear that a large proportion of genomic DNA in eukaryotic cells does not have any obvious functions (hence does not represent a "minimal set of genetic material"), and even functional segments are often dispensable (hence not "necessary and sufficient for the proper functioning of the organism"). Therefore, practically, the genome may be defined as "an entire set of genetic materials that is replicated and passed down to the next generation".

[Molecular] Phylogenetics and Phylogeny. Phylogenetics is a field of research concerned with the evolutionary trajectory and aims at the best possible reconstruction of the history of descent of organisms (i.e., phylogeny; how different but related species, extant and extinct, have emerged from their ancestors and form different branches/clades/phyla). Traditionally, phylogenetic analysis relied on phenotypical similarities (morphological, biochemical, etc.) and fossil records. Molecular phylogenetics is a newer division of phylogenetics which employs molecular sequence data (typically nucleotide sequence of DNA) to estimate (infer) the evolutionary history of molecules. The same methodology employed in molecular phylogenetics can be used to estimate organismal evolution if some criteria are met: e.g., (1) the genes have not been exchanged (shuffled) between species in the past, (2) the copies of the gene from different species are sufficiently similar to allow comparison but sufficiently different to provide necessary resolutions. However, sometimes these criteria are not met, often due to lateral gene transfers (which violates criteria 1) or too frequent/too infrequent mutations (which violates criteria 2). Thus, phylogenetic trees built from molecular sequences (gene trees) do not always reflect the "true" phylogenetic relationships of the organisms.

Three Domains of Life: Eukarya (eukaryotes), Bacteria (eubacteria), and Archaea (archaebacteria). All cellular life can be classified into three major groups ("domains"), i.e., eukarya (eukaryotes), bacteria (formerly called eubacteria), and archaea (formerly called archaebacteria). This classification is based on sequence similarities in ribosomal RNA, an essential component of protein biosynthesis machinery of all living cells. Eukarya are defined by the presence of a true nucleus, which envelopes genomic DNA, the genetic material to be replicated and passed down to the descendants. Bacteria and archaea lack a true nucleus (but nevertheless harbor genomic DNA within the cell) and are often collectively called prokaryotes. Bacteria and archaea are distinguished by several biochemical differences, such as the composition of cellular membrane lipids. Since these three types of cellular life are considered to have emerged a few billion years ago, estimating their precise relationship is still challenging and controversial.

Lectin. Proteins which recognize particular types of glycans are collectively called lectins. Antibodies which recognize glycans are excluded from this definition as are enzymes involved in the biosynthesis of glycans (which frequently recognize glycans/sugars as target substrates). The word lectin was originally used to describe plant proteins which recognize glycans/sugars, but the definition is now expanded to include such proteins from all sources.

Protozoa and Metazoa. Metazoa are multicellular animals ($\approx$ deuterostomes + protostomes). Protozoa are single-celled eukaryotic organisms which show similarity to animal cells rather than to plants or fungi. For example, amoebae, trypanosomes, plasmodia, and slime molds are all classified as protozoa. Because the classification "protozoa" includes so many different kinds of paraphyletic life forms, it is often considered not meaningful from the phylogenetic point of view. However, we used this word for convenience in this review because protozoan lineages did split from metazoa before the latter diverged into protostomes and deuterostomes.

GenBank. GenBank is the largest publicly accessible database of nucleotide and protein sequences maintained by National Center for Biotechnology Information (NCBI), a division of National Institute of Health (NIH). It cooperates with two other major sequence databases, i.e., the European Molecular Biology Laboratory (EMBL) and the DNA DataBank of Japan (DDBJ). NCBI's web site (http://www.ncbi.nlm.nih.gov/) provides gateways not only to GenBank but also to other useful literature (PubMed, OMIM), taxonomy, and protein structure databases.

Vertical vs Lateral/Horizontal [Gene] Transfer. The very basic (traditional) assumption of evolution is that genetic material is passed from the ancestor to descendant (vertical transfer) and exchange of genetic materials takes place only between the individuals belonging to the same (sexually reproducing) species. However, it is now clear that prokaryotes frequently engage in cross-species exchange of genetic materials (lateral transfer). Some evidence suggests that such lateral transfer is not limited to prokaryotes but also occurs between prokaryotes and single-cell eukaryotes (the direction is generally from prokaryote to eukaryote, rather than the other way around). This finding has important implications for the phylogeny of organisms: the universal phylogenetic tree of organisms may not be represented by a successively branching tree but rather by a complex network with some threads representing lateral transfers of individual genes.

Hemagglutinin (haemagglutinin). This word describes an entity which can cause agglutination of red blood cells ("hemocytes" in old terminology). It has been frequently used to describe viruses which cause hemagglutination. Lectins of viral origin which cause hemagglutination are traditionally called hemagglutinins rather than "viral lectins". However, it

should be noted that a hemagglutinin does not always recognize glycans on the red blood cells (the hemagglutinin may recognize proteins), and even when it recognizes glycans, sialic acid is not always the sugar being recognized.

Neuraminidase (sialidase). This is best defined as an enzyme which catalyzes hydrolysis of the α-ketoside linkage between sialic acids and proximal sugars. The two words are interchangeable.

Homolog, Paralog, and Ortholog. Consider a simple model where a pair of genes arose from a gene duplication, such that species A has two similar genes, Xa and Ya, and species B has two similar genes, Xb and Yb. Also assume that the common ancestor C (which no longer exists) of A and B had genes Xc and Yc which are closely related. The genes Xa and Xb, which are in different species but shared an immediate common ancestor Xc, are orthologs. The same applies to the pair [Ya,Yb]. The pairs [Xa,Xb], [Xa,Yb], [Xb,Ya], [Ya,Yb], which do not share immediate common ancestors but are nevertheless related by their ancestry, are called paralogs. All six pairs shown above are homologs.

Neighbor-Joining (distance-based method) and Maximum Parsimony. These two methods are frequently used in phylogenetic analysis to estimate (infer) the path of evolution. The former is a distance-based method which tries to estimate the evolutionary relationships of extant genes based on the matrix of pairwise distances or "sequence (dis)similarities" between all possible sequence pairs. The assumption is that the more similar two genes are, the more recently these two must have shared a common ancestor. The latter (maximum persimony) tries to emulate the path of evolution which explains current status of genes while minimizing the number of changes (mutations) necessary. There are some other methods available for phylogenetic analysis (tree reconstruction), notably the maximum likelihood method, based on detailed probablistic models of nucleotide changes.

Topology of the Phylogenetic Tree, Monophyly, and Paraphyly. The topology of a phylogenetic tree refers to how its branches are clustered together/split from each other, which in turn indicates how these genes/species are related to each other. A cluster of terminal branches connected by common ancestry which in turn is clearly set apart from the rest is called a monophyletic group (or clade) and the topology monophyly. For example, in the CMP-Neu5Ac synthetase clade of Figure 5A, the cluster [human,mouse,trout,fruit fly] is monophyletic, since these clusters include all branches descending from a common ancestor (the putative CMP-Neu5Ac synthetase gene in the ancestor of all metazoa). If we leave out the fruit fly from the above cluster (i.e., the cluster [human,mouse,trout]), it is still monophyletic, since it still includes all branches descending from a common ancestor (the putative CMP-Neu5Ac synthetase gene in the common ancestor of all vertebrates). On the other hand, if we leave out humans from the above cluster (i.e., the cluster [mouse,trout]), it is no longer monophyletic but it now becomes paraphyletic.

Root of Tree and Outgroup. The root of a tree, in short, represents the common ancestor of all genes (or organisms) represented in the phylogenetic tree. In traditional phylogenetics, a fossil which shows the presumed characteristics of a common ancestor for all organisms represented in the phylogenetic tree may serve as the root of the tree, if one can find such a fossil. However, in molecular phylogenetics, one cannot directly analyze such ancestral DNA sequences in most cases because such an ancestor is long dead and its DNA cannot be recovered. To define the root of a tree one needs to find an outgroup (evolutionary outlier), e.g., (1) a gene which shares common ancestry with the gene of interest or (2) an outlier within the group if one can find it a priori. In the case of phylogenetic analysis involving deep evolutionary time, neither is often possible. One way to define the root of a tree without using an outgroup is to find a midpoint from the tips of all terminal branches (midpoint rooting). However, this simple method explicitly assumes that the mutation rate has been constant on all the branches over time (in other words, "the molecular clock was in effect"), which is often an unrealistic assumption.

Different Modes of Cohabitation: Symbiosis, Commensalism, and Parasitism. When two different species live in close contact and one serves as the host to the other, there may be three different (but functionally continuous) modes of cohabitations. First, when the relationship benefits only the "guest" and is detrimental to the host, the relationship is defined as parasitism. Most pathogens belong to this category. Second, when the relationship has no obvious benefit to either organism but no harm either, such a relationship is called commensalism (from commensalis = "at table together" in Latin). Some of the bacteria (enterobacteria) living in vertebrate gastrointestinal tracts may belong to this group of commensal organisms. Finally, when the relationship is mutually beneficial (or at least beneficial to one of the partners and not detrimental to the other), such cohabitation is described as a symbiosis. For example, rhizobia (a group of bacteria) which inhabit the root nodules of leguminous plants and fix nitrogen are considered symbionts. Some intracellular organelles, such as mitochondria (found in higher eukaryotes) and chloroplasts (found in green plants), are considered as relics of ancient prokaryote/eukaryote symbiotic relationships.

X. Footnote

For all DNA database searches described in this review, we used the TBLASTN program[251] to search for homologs of the proteins of our interest in the nonredundant (NR) division of GenBank database at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov), unless otherwise stated. Amino acid sequences of the proteins, mostly those of the enzymes involved in the biosynthesis of sialic acids in human, mouse, and a bacterium *E. coli* strain K1, were used for such homology searches. Sequence matches with the Expect (E) value <1 were considered significant. The E value is a parameter that describes the number of hits one can "expect" to see

just by chance when searching a database of a particular size. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database one might expect to see 1 match with a similar score simply by chance.

When a member of a multigene family (e.g., mouse CMP-sialic acid transporter) was used as a template, many paralogs (e.g., mouse UDP-Gal transporter, human UDP-Gal transporter, human UDP-GlcNAc transporter, etc.) in addition to the putative orthologs in other organisms (e.g., human CMP-sialic acid transporter) were also found significantly similar to the template (query sequence). In such cases, "mutual best match" was employed as an additional criteria to define orthologs: for example, if gene X in organism A shows the highest similarity to gene Y in organism B (among all homologs in organism B) and Y shows the highest similarity to X (among all homologs in organism A), these two genes were considered orthologous.

Phylogenetic analyses were performed as follows: the amino acid sequences of an appropriate set of proteins were downloaded from GenBank database (for a comprehensive list of the proteins used for the construction of phylogenetic trees and their accession numbers, see Supporting Information). The amino acid sequences were aligned using ClustalW software (http://www2.ebi.ac.uk/clustalw/) and analyzed for the phylogenetic relationships using PAUP4.0 software package (Sinauer Associates). Distance-based matrixes of pairwise comparisons were used to construct phylogenetic trees by the neighbor-joining method.[206] No attempts were made to refine the alignments of the amino acid sequences prior to phylogenetic analysis. Sequence alignment files used for the construction of the phylogenetic trees are also available as Supporting Information.

XI. Supporting Information

Protein sequences used for the reconstruction of phylgenetic trees. This material is available free of charge via the Internet at http://pubs.acs.org.

XII. References

(1) Blix, G.; Gottschalk, A.; Klenk, E. *Nature* **1957**, *179*, 1088.
(2) Gottschalk, A. *The chemistry and biology of sialic acids and related substances*; University Press: Cambridge, 1960.
(3) Faillard, H. *TIBS* **1989**, *14*, 237.
(4) Nadano, D.; Iwasaki, M.; Endo, S.; Kitajima, K.; Inoue, S.; Inoue, Y. *J. Biol. Chem.* **1986**, *261*, 11550−11557.
(5) Inoue, S.; Kitajima, K.; Inoue, Y. *J. Biol. Chem.* **1996**, *271*, 24341−24344.
(6) Terada, T.; Kitazume, S.; Kitajima, K.; Inoue, S.; Ito, F.; Troy, F. A.; Inoue, Y. *J. Biol. Chem.* **1993**, *268*, 2640−2648.
(7) Angata, T.; Kitazume, S.; Terada, T.; Kitajima, K.; Inoue, S.; Troy, F. A., II.; Inoue, Y. *Glycoconjugate J.* **1994**, *11*, 493−499.
(8) Angata, T.; Nakata, D.; Matsuda, T.; Kitajima, K.; Troy, F. A., II *J. Biol. Chem.* **1999**, *274*, 22949−22956.
(9) Kawano, T.; Koyama, S.; Takematsu, H.; Kozutsumi, Y.; Kawasaki, H.; Kawashima, S.; Kawasaki, T.; Suzuki, A. *J. Biol. Chem.* **1995**, *270*, 16458−16463.
(10) Manzi, A. E.; Sjoberg, E. R.; Diaz, S.; Varki, A. *J. Biol. Chem.* **1990**, *265*, 13091−13103.
(11) Shukla, A. K.; Schroder, C.; Nohle, U.; Schauer, R. *Carbohydr. Res.* **1987**, *168*, 199−209.
(12) Luneberg, E.; Zetzmann, N.; Alber, D.; Knirel, Y. A.; Kooistra, O.; Zahringer, U.; Frosch, M. *Int. J. Med. Microbiol.* **2000**, *290*, 37−49.
(13) Schauer, R. *Sialic Acids: Chemistry, Metabolism and Function. Cell Biology Monographs*; Springer-Verlag: New York, 1982; Vol. 10.
(14) Rosenberg, A. *Biology of the sialic acids*; Plenum Press: New York, 1995.
(15) Schauer, R.; Kamerling, J. P. In *Glycoproteins II*; Montreuil, J., Vliegenthart, J. F. G., Schachter, H., Eds.; Elsevier: Amsterdam, 1997; Chapter 11, pp 243−402.
(16) *Essentials of Glycobiology*; Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., Marth, J., Eds.; Cold Spring Harbor Laboratory Press: New York, 1999.
(17) Woese, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8392−836.
(18) Philippe, H.; Forterre, P. *J. Mol. Evol.* **1999**, *49*, 509−523.
(19) Tsuji, S.; Datta, A. K.; Paulson, J. C. *Glycobiology* **1996**, *6*, V−VII.
(20) Iwasaki, M.; Inoue, S.; Kitajima, K.; Nomoto, H.; Inoue, Y. *Biochemistry* **1984**, *23*, 305−310.
(21) Muralikrishna, G.; Reuter, G.; Peter-Katalinic, J.; Egge, H.; Hanisch, F.-G.; Siebert, H.-C.; Schauer, R. *Carbohydr. Res.* **1992**, *236*, 321−326.
(22) Green, E. D.; Adelt, G.; Baenziger, J. U.; Wilson, S.; van Halbeek, H. *J. Biol. Chem.* **1988**, *263*, 18253−18268.
(23) Kitagawa, H.; Takaoka, M.; Nakada, H.; Fukui, S.; Funakoshi, I.; Kawasaki, T.; Tate, S.; Inagaki, F.; Yamashina, I. *J. Biochem. (Tokyo)* **1991**, *110*, 598−604.
(24) Kubo, H.; Irie, A.; Inagaki, F.; Hoshi, M. *J. Biochem. (Tokyo)* **1990**, *108*, 185−192.
(25) Takasaki, S.; Yamashita, K.; Suzuki, K.; Iwanaga, S.; Kobata, A. *J. Biol. Chem.* **1979**, *254*, 8548−8553.
(26) Slomiany, A.; Slomiany, B. L. *J. Biol. Chem.* **1978**, *253*, 7301−7306.
(27) Fukuda, M. N.; Dell, A.; Oates, J. E.; Fukuda, M. *J. Biol. Chem.* **1985**, *260*, 6623−6631.
(28) Kitazume, S.; Kitajima, K.; Inoue, S.; Troy, F. A., II.; Cho, J.-W.; Lennarz, W. J.; Inoue, Y. *J. Biol. Chem.* **1994**, *269*, 22712−22718.
(29) Maes, E.; Wieruszeski, J.-M.; Plancke, Y.; Strecker, G. *FEBS Lett.* **1995**, *358*, 205−210.
(30) Zanetta, J. P.; Pons, A.; Iwersen, M.; Mariller, C.; Leroy, Y.; Timmerman, P.; Schauer, R. *Glycobiology* **2001**, *11*, 663−676.
(31) Mitsuoka, C.; Ohmori, K.; Kimura, N.; Kanamori, A.; Komba, S.; Ishida, H.; Kiso, M.; Kannagi, R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1597−1602.
(32) Cebo, C.; Dambrouck, T.; Maes, E.; Laden, C.; Strecker, G.; Michalski, J. C.; Zanetta, J. P. *J. Biol. Chem.* **2001**, *276*, 5685−5691.
(33) Inoue, S. *Biochim. Biophys. Acta* **1965**, *101*, 16−25.
(34) Warren, L. *Comp. Biochem. Physiol.* **1963**, *10*, 153−171.
(35) Davidson, D. J.; Castellino, F. J. *J. Clin. Invest.* **1993**, *92*, 249−254.
(36) Marchal, I.; Jarvis, D. L.; Cacan, R.; Verbert, A. *Biol. Chem.* **2001**, *382*, 151−159.
(37) Altmann, F.; Staudacher, E.; Wilson, I. B. H.; März, L. *Glycoconjugate J.* **1999**, *16*, 109−123.
(38) Roth, J.; Kempf, A.; Reuter, G.; Schauer, R.; Gehring, W. J. *Science* **1992**, *256*, 673−675.
(39) Malykh, Y. N.; Krisch, B.; Gerardy-Schahn, R.; Lapina, E. B.; Shaw, L.; Schauer, R. *Glycoconjugate J.* **1999**, *16*, 731−739.
(40) Saito, M.; Kitamura, H.; Sugiyama, K. *Biochim. Biophys. Acta Bio-Membr.* **2001**, *1511*, 271−280.
(41) Alviano, C. S.; Travassos, L. R.; Schauer, R. *Glycoconjugate J.* **1999**, *16*, 545−554.
(42) Soares, R. M. A.; Soares, R. M. D.; Alviano, D. S.; Angluster, J.; Alviano, C. S.; Travassos, L. R. *Biochim. Biophys. Acta Gen. Subj.* **2000**, *1474*, 262−268.
(43) Rodrigues, M. L.; Rozental, S.; Couceiro, J. N.; Angluster, J.; Alviano, C. S.; Travassos, L. R. *Infect. Immun.* **1997**, *65*, 4937−4942.
(44) Wasylnka, J. A.; Simmer, M. I.; Moore, M. M. *Microbiology* **2001**, *147*, 869−877.
(45) Alviano, C. S.; Pereira, M. E. A.; Souza, W.; Oda, L. M.; Travassos, L. R. *FEMS Microbiol. Lett.* **1982**, *15*, 223−227.
(46) Matsumoto, S.; Ikura, K.; Ueda, M.; Sasaki, R. *Plant Mol. Biol.* **1995**, *27*, 1163−1172.
(47) Lerouge, P.; Cabanes-Macheteau, M.; Rayon, C.; Fischette-Lainé, A. C.; Gomord, V.; Faye, L. *Plant Mol. Biol.* **1998**, *38*, 31−48.
(48) Bourbouze, R.; Akiki, C.; Chardon-Loriaux, I.; Percheron, F. *Carbohydr. Res.* **1982**, *106*, 21−30.
(49) Schauer, R.; Reuter, G.; Muhlpfordt, H.; Andrade, A. F.; Pereira, M. E. *Hoppe-Seyler's Z. Physiol. Chem.* **1983**, *364*, 1053−1057.
(50) Cross, G. A. M.; Takle, G. B. *Annu. Rev. Microbiol.* **1993**, *47*, 385−411.
(51) Belen Carrillo, M.; Gao, W.; Herrera, M.; Alroy, J.; Moore, J. B.; Beverley, S. M.; Pereira, M. A. *Infect. Immun.* **2000**, *68*, 2728−2734.
(52) Yoshida, M.; Fuse, G.; Matsui, T.; Ouchi, S. *Biochem. Biophys. Res. Commun.* **1992**, *188*, 794−798.
(53) Matta, M. A.; Aleksitch, V.; Angluster, J.; Alviano, C. S.; De Souza, W.; Andrade, A. F.; Esteves, M. J. *Parasitol. Res.* **1995**, *81*, 426−433.

(54) Watarai, S.; Sugimoto, C.; Hosotani-Kaihara, K.; Kobayashi, K.; Onuma, M.; Lee, J. T.; Kushi, Y.; Handa, S.; Yasuda, T. *J. Vet. Med. Sci.* **1996**, *58*, 1099−1105.
(55) Avron, B.; Chayen, A.; Stolarsky, T.; Schauer, R.; Reuter, G.; Mirelman, D. *Mol. Biochem. Parasitol.* **1987**, *25*, 257−266.
(56) Sorice, M.; Griggi, T.; Nicodemo, G.; Garofalo, T.; Marangi, M.; Sanguigni, S.; Becker, S. I.; Mirelman, D. *Parasite Immunol.* **1996**, *18*, 133−137.
(57) Hood, D. W.; Cox, A. D.; Gilbert, M.; Makepeace, K.; Walsh, S.; Deadman, M. E.; Cody, A.; Martin, A.; Månsson, M.; Schweda, E. K. H.; Brisson, J. R.; Richards, J. C.; Moxon, E. R.; Wakarchuk, W. W. *Mol. Microbiol.* **2001**, *39*, 341−350.
(58) Gilbert, M.; Brisson, J. R.; Karwaski, M. F.; Michniewicz, J.; Cunningham, A. M.; Wu, Y. Y.; Young, N. M.; Wakarchuk, W. W. *J. Biol. Chem.* **2000**, *275*, 3896−3906.
(59) Wakarchuk, W. W.; Watson, D.; St Michael, F.; Li, J. J.; Wu, Y. Y.; Brisson, J. R.; Young, N. M.; Gilbert, M. *J. Biol. Chem.* **2001**, *276*, 12785−12790.
(60) Reuter, G.; Gabius, H. J. *Biol. Chem. Hoppe Seyler* **1996**, *377*, 325−342.
(61) Knirel, Y. A.; Kocharova, N. A.; Shashkov, A. S.; Kochetkov, N. K.; Mamontova, V. A.; Soloveva, T. F. *Carbohydr. Res.* **1989**, *188*, 145−155.
(62) Gil-Serrano, A. M.; Rodríguez-Carvajal, M. A.; Tejero-Mateo, P.; Espartero, J. L.; Thomas-Oates, J.; Ruiz-Sainz, J. E.; Buendía-Clavería, A. M. *Biochem. J.* **1998**, *334*, 585−594.
(63) Ochman, H.; Lawrence, J. G.; Groisman, E. A. *Nature* **2001**, *405*, 299−304.
(64) Eisen, J. A. *Curr. Opin. Genet. Dev.* **2000**, *10*, 606−611.
(65) Troy, F. A., II *Glycobiology* **1992**, *2*, 5−23.
(66) Glode, M. P.; Sutton, A.; Robbins, J. B.; McCracken, G. H.; Gotschlich, E. C.; Kaijser, B.; Hanson, L. A. *J. Infect. Dis.* **1977**, *136 Suppl*, S93−S97.
(67) Yuki, N. *Cell. Mol. Life Sci.* **2000**, *57*, 527−533.
(68) Schauer, R. *Adv. Carbohydr. Chem. Biochem.* **1982**, *40*, 131−234.
(69) Varki, A. *Glycobiology* **1992**, *2*, 25−40.
(70) Krauss, J. H.; Himmelspach, K.; Reuter, G.; Schauer, R.; Mayer, H. *Eur. J. Biochem.* **1992**, *204*, 217−223.
(71) Defives, C.; Bouslamti, R.; Derieux, J. C.; Kol, O.; Fournet, B. *FEMS Microbiol. Lett.* **1989**, *57*, 203−208.
(72) Zähringer, U.; Knirel, Y. A.; Lindner, B.; Helbig, J. H.; Sonesson, A.; Marre, R.; Rietschel, E. Th. In *Bacterial endotoxins: lipopolysaccharides from genes to therapy*; Levin, C. R., Alving, R. S., Munford, R. S., Redl, H., Eds.; Wiley-Liss: New York, 1995; pp 113−139.
(73) Tsvetkov, Y. E.; Shashkov, A. S.; Knirel, Y. A.; Zähringer, U. *Carbohydr. Res.* **2001**, *331*, 233−237.
(74) Gil-Serrano, A. M.; Rodriguez-Carvajal, M. A.; Tejero-Mateo, P.; Espartero, J. L.; Menendez, M.; Corzo, J.; Ruiz-Sainz, J. E.; BuendiA-Claveria, A. M. *Biochem. J.* **1999**, *342 Pt 3*, 527−535.
(75) Muldoon, J.; Shashkov, A. S.; Senchenkova, S. N.; Tomshich, S. V.; Komandrova, N. A.; Romanenko, L. A.; Knirel, Y. A.; Savage, A. V. *Carbohydr. Res.* **2001**, *330*, 231−239.
(76) Knirel, Y. A.; Grosskurth, H.; Helbig, J. H.; Zahringer, U. *Carbohydr. Res.* **1995**, *279*, 215−226.
(77) Defives, C.; Werquin, M.; Mary, P.; Hornez, J. P. *Curr. Microbiol.* **1996**, *33*, 371−376.
(78) Reuhs, B. L.; Geller, D. P.; Kim, J. S.; Fox, J. E.; Kolli, V. S.; Pueppke, S. G. *Appl. Environ. Microbiol.* **1998**, *64*, 4930−4938.
(79) Armstrong, P. B.; Swarnakar, S.; Srimal, S.; Misquith, S.; Hahn, E. A.; Aimes, R. T.; Quigley, J. P. *J. Biol. Chem.* **1996**, *271*, 14717−14721.
(80) Tunkijjanukij, S.; Giaever, H.; Chin, C. C. Q.; Olafsen, J. A. *Comp. Biochem. Physiol. [B]* **1998**, *119B*, 705−713.
(81) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. *Nature* **1981**, *289*, 366−373.
(82) Sauter, N. K.; Glick, G. D.; Crowther, R. L.; Park, S. J.; Eisen, M. B.; Skehel, J. J.; Knowles, J. R.; Wiley: D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 324−328.
(83) Rosenthal, P. B.; Zhang, X. D.; Formanowski, F.; Fitz, W.; Wong, C. H.; Meier-Ewert, H.; Skehel, J. J.; Wiley, D. C. *Nature* **1998**, *396*, 92−96.
(84) Crennell, S.; Takimoto, T.; Portner, A.; Taylor, G. *Nat. Struct. Biol.* **2000**, *7*, 1068−1074.
(85) Stehle, T.; Yan, Y.; Benjamin, T. L.; Harrison, S. C. *Nature* **1994**, *369*, 160−163.
(86) Suzuki, Y.; Ito, T.; Suzuki, T.; Holland, R. E., Jr.; Chambers, T. M.; Kiso, M.; Ishida, H.; Kawaoka, Y. *J. Virol.* **2000**, *74*, 11825−11831.
(87) Webster, R. G.; Bean, W. J.; Gorman, O. T.; Chambers, T. M.; Kawaoka, Y. *Microbiol. Rev.* **1992**, *56*, 152−179.
(88) Ito, T.; Couceiro, J. N.; Kelm, S.; Baum, L. G.; Krauss, S.; Castrucci, M. R.; Donatelli, I.; Kida, H.; Paulson, J. C.; Webster, R. G.; Kawaoka, Y. *J. Virol.* **1998**, *72*, 7367−7373.
(89) Rogers, G. N.; Herrler, G.; Paulson, J. C.; Klenk, H. D. *J. Biol. Chem.* **1986**, *261*, 5947−5951.
(90) Herrler, G.; Rott, R.; Klenk, H. D.; Muller, H. P.; Shukla, A. K.; Schauer, R. *EMBO J.* **1985**, *4*, 1503−1506.
(91) Muchmore, E.; Varki, A. *Science* **1987**, *236*, 1293−1295.
(92) Klein, A.; Krishna, M.; Varki, N. M.; Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7782−7786.
(93) Sjoberg, E. R.; Powell, L. D.; Klein, A.; Varki, A. *J. Cell Biol.* **1994**, *126*, 549−562.
(94) Shi, W. X.; Chammas, R.; Varki, N. M.; Powell, L.; Varki, A. *J. Biol. Chem.* **1996**, *271*, 31526−31532.
(95) Regl, G.; Kaser, A.; Iwersen, M.; Schmid, H.; Kohla, G.; Strobl, B.; Vilas, U.; Schauer, R.; Vlasak, R. *J. Virol.* **1999**, *73*, 4721−4727.
(96) Suzuki, T.; Portner, A.; Scroggs, R. A.; Uchikawa, M.; Koyama, N.; Matsuo, K.; Suzuki, Y.; Takimoto, T. *J. Virol.* **2001**, *75*, 4604−4613.
(97) Baum, L. G.; Paulson, J. C. *Virology* **1991**, *180*, 10−15.
(98) Mitnaul, L. J.; Matrosovich, M. N.; Castrucci, M. R.; Tuzikov, A. B.; Bovin, N. V.; Kobasa, D.; Kawaoka, Y. *J. Virol.* **2000**, *74*, 6015−6020.
(99) Jaramillo, L.; Diaz, F.; Hernandez, P.; Debray, H.; Trigo, F.; Mendoza, G.; Zenteno, E. *Glycobiology* **2000**, *10*, 31−37.
(100) Rousset, E.; Harel, J.; Dubreuil, J. D. *Infect. Immun.* **1998**, *66*, 5650−5658.
(101) Ofek, I.; Sharon, N. *Curr. Top. Microbiol. Immunol.* **1990**, *151*, 91−114.
(102) Miller-Podraza, H.; Milh, M. A.; Bergström, J.; Karlsson, K. A. *Glycoconjugate J.* **1996**, *13*, 453−460.
(103) Miller-Podraza, H.; Bergstrom, J.; Milh, M. A.; Karlsson, K. A. *Glycoconjugate J.* **1997**, *14*, 467−471.
(104) Ono, E.; Abe, K.; Nakazawa, M.; Naiki, M. *Infect. Immun.* **1989**, *57*, 907−911.
(105) Smit, H.; Gaastra, W.; Kamerling, J. P.; Vliegenthart, J. F. G.; de Graaf, F. K. *Infect. Immun.* **1984**, *46*, 578−584.
(106) Yuyama, Y.; Yoshimatsu, K.; Ono, E.; Saito, M.; Naiki, M. *J. Biochem. (Tokyo)* **1993**, *113*, 488−492.
(107) Richards, R. L.; Moss, J.; Alving, C. R.; Fishman, P. H.; Brady, R. O. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 1673−1676.
(108) Orlandi, P. A.; Klotz, F. W.; Haynes, J. D. *J. Cell Biol.* **1992**, *116*, 901−909.
(109) Dolan, S. A.; Miller, L. H.; Wellems, T. E. *J. Clin. Invest.* **1990**, *86*, 618−624.
(110) Reed, M. B.; Caruana, S. R.; Batchelor, A. H.; Thompson, J. K.; Crabb, B. S.; Cowman, A. F. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7509−7514.
(111) Okoyeh, J. N.; Pillai, C. R.; Chitnis, C. E. *Infect. Immun.* **1999**, *67*, 5784−5791.
(112) Mayer, D. C. G.; Kaneko, O.; Hudson-Taylor, D. E.; Reid, M. E.; Miller, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5222−5227.
(113) Babál, P.; Pindak, F. F.; Russell, L. C.; Gardner, W. A., Jr. *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1428*, 106−116.
(114) Feingold, C.; Mirelman, D.; Lotan, D.; Lotan, R. *Cancer Lett.* **1984**, *24*, 263−271.
(115) Ram, S.; Sharma, A. K.; Simpson, S. D.; Gulati, S.; McQuillen, D. P.; Pangburn, M. K.; Rice, P. A. *J. Exp. Med.* **1998**, *187*, 743−752.
(116) Kazatchkine, M. D.; Fearon, D. T.; Austen, K. F. *J. Immunol.* **1979**, *122*, 75−81.
(117) Bevilacqua, M.; Butcher, E.; Furie, B.; Gallatin, M.; Gimbrone, M.; Harlan, J.; Kishimoto, K.; Lasky, L.; McEver, R.; Paulson, J.; Rosen, S.; Seed, B.; Siegelman, M.; Springer, T.; Stoolman, L.; Tedder, T.; Varki, A.; Wagner, D.; Weissman, I.; Zimmerman, G. *Cell* **1991**, *67*, 233.
(118) Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7390−7397.
(119) Lasky, L. A. *Annu. Rev. Biochem.* **1995**, *64*, 113−139.
(120) Kim, Y. J.; Borsig, L.; Varki, N. M.; Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9325−9330.
(121) Wilkins, P. P.; Moore, K. L.; McEver, R. P.; Cummings, R. D. *J. Biol. Chem.* **1995**, *270*, 22677−22680.
(122) Somers, W. S.; Tang, J.; Shaw, G. D.; Camphausen, R. T. *Cell* **2000**, *103*, 467−479.
(123) Norgard-Sumnicht, K. E.; Varki, N. M.; Varki, A. *Science* **1993**, *261*, 480−483.
(124) Frenette, P. S.; Wagner, D. D. *Thromb. Haemost.* **1997**, *78*, 60−64.
(125) Maly, P.; Thall, A. D.; Petryniak, B.; Rogers, G. E.; Smith, P. L.; Marks, R. M.; Kelly, R. J.; Gersten, K. M.; Cheng, G. Y.; Saunders: T. L.; Camper, S. A.; Camphausen, R. T.; Sullivan, F. X.; Isogai, Y.; Hindsgaul, O.; Von Andrian, U. H.; Lowe, J. B. *Cell* **1996**, *86*, 643−653.
(126) Borsig, L.; Wong, R.; Feramisco, J.; Nadeau, D. R.; Varki, N. M.; Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3352−3357.
(127) Leshko-Lindsay, L.; Corces, V. G. *Development* **1997**, *124*, 169−180.
(128) Crocker, P. R.; Clark, E. A.; Filbin, M.; Gordon, S.; Jones, Y.; Kehrl, J. H.; Kelm, S.; Le Douarin, N.; Powell, L.; Roder, J.; Schnaar, R. L.; Sgroi, D. C.; Stamenkovic, K.; Schauer, R.; Schachner, M.; Van den Berg, T. K.; Van der Merwe, P. A.; Watt, S. M.; Varki, A. *Glycobiology* **1998**, *8*, v.
(129) Crocker, P. R.; Varki, A. *Trends Immunol.* **2001**, *22*, 337−342.
(130) Angata, T.; Hingorani, R.; Varki, N. M.; Varki, A. *J. Biol. Chem.* **2001**, *276*, 45128−45136.

(131) O'Keefe, T. L.; Williams, G. T.; Davies, S. L.; Neuberger, M. S. *Science* **1996**, *274*, 798−801.
(132) Otipoby, K. L.; Andersson, K. B.; Draves, K. E.; Klaus, S. J.; Farr, A. G.; Kerner, J. D.; Perlmutter, R. M.; Law, C. L.; Clark, E. A. *Nature* **1996**, *384*, 634−637.
(133) Sato, S.; Miller, A. S.; Inaoki, M.; Bock, C. B.; Jansen, P. J.; Tang, M. L. K.; Tedder, T. F. *Immunity* **1996**, *5*, 551−562.
(134) Nitschke, L.; Carsetti, R.; Ocker, B.; Köhler, G.; Lamers, M. C. *Curr. Biol.* **1997**, *7*, 133−143.
(135) Schachner, M.; Bartsch, U. *Glia* **2000**, *29*, 154−165.
(136) May, A. P.; Robinson, R. C.; Vinson, M.; Crocker, P. R.; Jones, E. Y. *Mol. Cell* **1998**, *1*, 719−728.
(137) Vinson, M.; Van der Merwe, P. A.; Kelm, S.; May, A.; Jones, E. Y.; Crocker, P. R. *J. Biol. Chem.* **1996**, *271*, 9267−9272.
(138) Van der Merwe, P. A.; Crocker, P. R.; Vinson, M.; Barclay, A. N.; Schauer, R.; Kelm, S. *J. Biol. Chem.* **1996**, *271*, 9273−9280.
(139) Tang, S.; Shen, Y. J.; DeBellard, M. E.; Mukhopadhyay, G.; Salzer, J. L.; Crocker, P. R.; Filbin, M. T. *J. Cell Biol.* **1997**, *138*, 1355−1366.
(140) Angata, T.; Varki, N. M.; Varki, A. *J. Biol. Chem.* **2001**, *276*, 40282−40287.
(141) Dulac, C.; Tropak, M. B.; Cameron-Curry, P.; Rossier, J.; Marshak, D. R.; Roder, J.; Le Douarin, N. M. *Neuron.* **1992**, *8*, 323−334.
(142) Angata, T.; Varki, A. *J. Biol. Chem.* **2000**, *275*, 22127−22135.
(143) Scholler, N.; Hayden-Ledbetter, M.; Hellström, K. E.; Hellström, I.; Ledbetter, J. A. *J. Immunol.* **2001**, *166*, 3865−3872.
(144) Kleene, R.; Yang, H.; Kutsche, M.; Schachner, M. *J. Biol. Chem.* **2001**, *276*, 21656−21663.
(145) Chiba, A.; Matsumura, K.; Yamada, H.; Inazu, T.; Shimizu, T.; Kusunoki, S.; Kanazawa, I.; Kobata, A.; Endo, T. *J. Biol. Chem.* **1997**, *272*, 2156−2162.
(146) Iglesias, M. M.; Cymes, G. D.; Wolfenstein-Todel, C. *Glycoconjugate J.* **1996**, *13*, 967−976.
(147) Zeng, F. Y.; Gabius, H. J. *Biochem. Int.* **1992**, *26*, 17−24.
(148) Zeng, F. Y.; Gabius, H. J. *Arch. Biochem. Biophys.* **1991**, *289*, 137−144.
(149) Brinkman-Van der Linden, E. C. M.; Varki, A. *J. Biol. Chem.* **2000**, *275*, 8625−8632.
(150) Vacquier, V. D. *Science* **1998**, *281*, 1995−1998.
(151) Nagai, Y.; Hoshi, M. *Biochim. Biophys. Acta* **1975**, *388*, 146−151.
(152) Gagneux, P.; Varki, A. *Glycobiology* **1999**, *9*, 747−755.
(153) Van Valen, L. *Nature* **1974**, *252*, 298−300.
(154) Muchmore, E.; Varki, N.; Fukuda, M.; Varki, A. *FASEB J.* **1987**, *1*, 229−235.
(155) Sharma, V.; Chatterjee, M.; Mandal, C.; Sen, S.; Basu, D. *Am. J. Trop. Med. Hyg.* **1998**, *58*, 551−554.
(156) Drzeniek, R. *Histochem. J.* **1973**, *5*, 271−290.
(157) Corfield, A. P.; Veh, R. W.; Wember, M.; Michalski, J. C.; Schauer, R. *Biochem. J.* **1981**, *197*, 293−299.
(158) Corfield, A. P.; Sander Wewer, M.; Veh, R. W.; Wember, M.; Schauer, R. *Biol. Chem. Hoppe Seyler* **1986**, *367*, 433−439.
(159) Varki, A. *FASEB J.* **1997**, *11*, 248−255.
(160) Muchmore, E. A.; Diaz, S.; Varki, A. *Am. J. Phys. Anthropol.* **1998**, *107*, 187−198.
(161) Chou, H. H.; Takematsu, H.; Diaz, S.; Iber, J.; Nickerson, E.; Wright, K. L.; Muchmore, E. A.; Nelson, D. L.; Warren, S. T.; Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11751−11756.
(162) Higashi, H.; Naiki, M.; Matuo, S.; Okouchi, K. *Biochem. Biophys. Res. Commun.* **1977**, *79*, 388−395.
(163) Kwiatkowski, B.; Boschek, B.; Thiele, H.; Stirm, S. *J. Virol.* **1983**, *45*, 367−374.
(164) Wollin, R.; Stocker, B. A.; Lindberg, A. A. *J. Bacteriol.* **1987**, *169*, 1003−1009.
(165) Rosenberg, A.; Schengrund, C.-L. *Biological Roles of Sialic Acids*; Plenum Press: New York and London, 1976.
(166) Bliss, J. M.; Silver, R. P. *Mol. Microbiol.* **1996**, *21*, 221−231.
(167) Edwards, U.; Müller, A.; Hammerschmidt, S.; Gerardy-Schahn, R.; Frosch, M. *Mol. Microbiol.* **1994**, *14*, 141−149.
(168) Ganguli, S.; Zapata, G.; Wallis, T.; Reid, C.; Boulnois, G.; Vann, W. F.; Roberts, I. S. *J. Bacteriol.* **1994**, *176*, 4583−4589.
(169) Hinderlich, S.; Stäsche, R.; Zeitler, R.; Reutter, W. *J. Biol. Chem.* **1997**, *272*, 24313−24318.
(170) Stäsche, R.; Hinderlich, S.; Weise, C.; Effertz, K.; Lucka, L.; Moormann, P.; Reutter, W. *J. Biol. Chem.* **1997**, *272*, 24319−24324.
(171) Hinderlich, S.; Berger, M.; Keppler, O. T.; Pawlita, M.; Reutter, W. *Biol. Chem.* **2001**, *382*, 291−297.
(172) Zapata, G.; Crowley, J. M.; Vann, W. F. *J. Bacteriol.* **1992**, *174*, 315−319.
(173) Petersen, M.; Fessner, W.; Frosch, M.; Luneberg, E. *FEMS Microbiol. Lett.* **2000**, *184*, 161−164.
(174) Daines, D. A.; Wright, L. F.; Chaffin, D. O.; Rubens, C. E.; Silver, R. P. *FEMS Microbiol. Lett.* **2000**, *189*, 281−284.
(175) Campbell, R. E.; Mosimann, S. C.; Tanner, M. E.; Strynadka, N. C. J. *Biochemistry* **2000**, *39*, 14993−15001.
(176) Lawrence, S. M.; Huddleston, K. A.; Pitts, L. R.; Nguyen, N.; Lee, Y. C.; Vann, W. F.; Coleman, T. A.; Betenbaugh, M. J. *J. Biol. Chem.* **2000**, *275*, 17869−17877.
(177) Nakata, D.; Close, B. E.; Colley, K. J.; Matsuda, T.; Kitajima, K. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 642−648.
(178) Vann, W. F.; Tavarez, J. J.; Crowley, J.; Vimr, E.; Silver, R. P. *Glycobiology* **1997**, *7*, 697−701.
(179) Lawrence, M. C.; Barbosa, J. A. R. G.; Smith, B. J.; Hall, N. E.; Pilling, P. A.; Ooi, H. C.; Marcuccio, S. M. *J. Mol. Biol.* **1997**, *266*, 381−399.
(180) Andersson, J. O.; Doolittle, W. F.; Nesbo, C. L. *Science* **2001**, *292*, 1848−1850.
(181) Zapata, G.; Vann, W. F.; Aaronson, W.; Lewis, M. S.; Moos, M. *J. Biol. Chem.* **1989**, *264*, 14769−14774.
(182) Edwards, U.; Frosch, M. *FEMS Microbiol. Lett.* **1992**, *96*, 161−166.
(183) Münster, A. K.; Eckhardt, M.; Potvin, B.; Mühlenhoff, M.; Stanley, P.; Gerardy-Schahn, R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9140−9145.
(184) Eckhardt, M.; Mühlenhoff, V.; Bethe, A.; Gerardy-Schahn, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7572−7576.
(185) Tsuji, S. *J. Biochem. (Tokyo)* **1996**, *120*, 1−13.
(186) Harduin-Lepers, A.; Recchi, M. A.; Delannoy, P. *Glycobiology* **1995**, *5*, 741−758.
(187) Livingston, B. D.; Paulson, J. C. *J. Biol. Chem.* **1993**, *268*, 11504−11507.
(188) Higa, H. H.; Paulson, J. C. *J. Biol. Chem.* **1985**, *260*, 8838−8849.
(189) Angata, T.; Matsuda, T.; Kitajima, K. *Glycobiology* **1998**, *8*, 277−284.
(190) Steenbergen, S. M.; Wrona, T. J.; Vimr, E. R. *J. Bacteriol.* **1992**, *174*, 1099−1108.
(191) Yamamoto, T.; Nakashizuka, M.; Kodama, H.; Kajihara, Y.; Terada, I. *J. Biochem. (Tokyo)* **1996**, *120*, 104−110.
(192) Gilbert, M.; Watson, D. C.; Cunningham, A. M.; Jennings, M. P.; Young, N. M.; Wakarchuk, W. W. *J. Biol. Chem.* **1996**, *271*, 28271−28276.
(193) Bozue, J. A.; Tullius, M. V.; Wang, J.; Gibson, B. W.; Munson, R. S., Jr. *J. Biol. Chem.* **1999**, *274*, 4106−4114.
(194) Meier-Dieter, U.; Starman, R.; Barr, K.; Mayer, H.; Rick, P. D. *J. Biol. Chem.* **1990**, *265*, 13490−13497.
(195) Vimr, E.; Lichtensteiger, C.; Steenbergen, S. *Mol. Microbiol.* **2000**, *36*, 1113−1123.
(196) Nairn, C. A.; Cole, J. A.; Patel, P. V.; Parsons, N. J.; Fox, J. E.; Smith, H. *J. Gen. Microbiol.* **1988**, *134*, 3295−3306.
(197) Bult, C. J.; White, O.; Olsen, G. J.; Zhou, L.; Fleischmann, R. D.; Sutton, G. G.; Blake, J. A.; FitzGerald, L. M.; Clayton, R. A.; Gocayne, J. D.; Kerlavage, A. R.; Dougherty, B. A.; Tomb, J. F.; Adams, M. D.; Reich, C. I.; Overbeek, R.; Kirkness, E. F.; Weinstock, K. G.; Merrick, J. M.; Glodek, A.; Scott, J. L.; Geoghagen, N. S.; Venter, J. C. *Science* **1996**, *273*, 1058−1073.
(198) Knirel, Y. A.; Rietschel, E. T.; Marre, R.; Zahringer, U. *Eur. J. Biochem.* **1994**, *221*, 239−245.
(199) Adams, M. D.; Celniker, S. E.; Holt, R. A.; Evans, C. A.; Gocayne, J. D.; Amanatides, P. G.; Scherer, S. E.; Li, P. W.; Hoskins, R. A.; Galle, R. F.; George, R. A.; Lewis, S. E.; Richards, S.; Ashburner, M.; Henderson, S. N.; Sutton, G. G.; Wortman, J. R.; Yandell, M. D.; Zhang, Q.; Chen, L. X.; Brandon, R. C.; Rogers, Y. H.; Blazej, R. G.; Champe, M.; Pfeiffer, B. D.; Wan, K. H.; Doyle, C.; Baxter, E. G.; Helt, G.; Nelson, C. R.; Gabor, G. L.; Abril, J. F.; Agbayani, A.; An, H. J.; Andrews-Pfannkoch, C.; Baldwin, D.; Ballew, R. M.; Basu, A.; Baxendale, J.; Bayraktaroglu, L.; Beasley, E. M.; Beeson, K. Y.; Benos, P. V.; Berman, B. P.; Bhandari, D.; Bolshakov, S.; Borkova, D.; Botchan, M. R.; Bouck, J.; Brokstein, P.; Brottier, P.; Burtis, K. C.; Busam, D. A.; Butler, H.; Cadieu, E.; Center, A.; Chandra, I.; Cherry, J. M.; Cawley, S.; Dahlke, C.; Davenport, L. B.; Davies, P.; de Pablos, B.; Delcher, A.; Deng, Z.; Mays, A. D.; Dew, I.; Dietz, S. M.; Dodson, K.; Doup, L. E.; Downes, M.; Dugan-Rocha, S.; Dunkov, B. C.; Dunn, P.; Durbin, K. J.; Evangelista, C. C.; Ferraz, C.; Ferriera, S.; Fleischmann, W.; Fosler, C.; Gabrielian, A. E.; Garg, N. S.; Gelbart, W. M.; Glasser, K.; Glodek, A.; Gong, F.; Gorrell, J. H.; Gu, Z.; Guan, P.; Harris, M.; Harris, N. L.; Harvey, D.; Heiman, T. J.; Hernandez, J. R.; Houck, J.; Hostin, D.; Houston, K. A.; Howland, T. J.; Wei, M. H.; Ibegwam, C.; Jalali, M.; Kalush, F.; Karpen, G. H.; Ke, Z.; Kennison, J. A.; Ketchum, K. A.; Kimmel, B. E.; Kodira, C. D.; Kraft, C.; Kravitz, S.; Kulp, D.; Lai, Z.; Lasko, P.; Lei, Y.; Levitsky, A. A.; Li, J.; Li, Z.; Liang, Y.; Lin, X.; Liu, X.; Mattei, B.; McIntosh, T. C.; McLeod, M. P.; McPherson, D.; Merkulov, G.; Milshina, N. V.; Mobarry, C.; Morris, J.; Moshrefi, A.; Mount, S. M.; Moy, M.; Murphy, B.; Murphy, L.; Muzny, D. M.; Nelson, D. L.; Nelson, D. R.; Nelson, K. A.; Nixon, K.; Nusskern, D. R.; Pacleb, J. M.; Palazzolo, M.; Pittman, G. S.; Pan, S.; Pollard, J.; Puri, V.; Reese, M. G.; Reinert, K.; Remington, K.; Saunders, R. D.; Scheeler, F.; Shen, H.; Shue, B. C.; Siden-Kiamos, I.; Simpson, M.; Skupski, M. P.; Smith, T.; Spier, E.; Spradling, A. C.; Stapleton, M.; Strong, R.; Sun, E.; Svirskas, R.; Tector, C.; Turner, R.; Venter, E.; Wang, A. H.; Wang, X.; Wang, Z. Y.; Wassarman, D. A.; Weinstock, G. M.; Weissenbach, J.; Williams, S. M.; Woodage,

T.; Worley, K. C.; Wu, D.; Yang, S.; Yao, Q. A.; Ye, J.; Yeh, R. F.; Zaveri, J. S.; Zhan, M.; Zhang, G.; Zhao, Q.; Zheng, L.; Zheng, X. H.; Zhong, F. N.; Zhong, W.; Zhou, X.; Zhu, S.; Zhu, X.; Smith, H. O.; Gibbs, R. A.; Myers, E. W.; Rubin, G. M.; Venter, J. C. *Science* **2000**, *287*, 2185−2195.
(200) Maru, I.; Ohta, Y.; Murata, K.; Tsukada, Y. *J. Biol. Chem.* **1996**, *271*, 16294−16299.
(201) The *C.elegans* Sequencing Consortium. *Science* **1998**, *282*, 2012−2018.
(202) Bacic, A.; Kahane, I.; Zuckerman, B. M. *Exp. Parasitol.* **1990**, *71*, 483−488.
(203) Dotson, G. D.; Nanjappan, P.; Reily, M. D.; Woodard, R. W. *Biochemistry* **1993**, *32*, 12392−12397.
(204) Shumilin, I. A.; Kretsinger, R. H.; Bauerle, R. H. *Struct. Fold. Des.* **1999**, *7*, 865−875.
(205) Radaev, S.; Dastidar, P.; Patel, M.; Woodard, R. W.; Gatti, D. L. *J. Biol. Chem.* **2000**, *275*, 9476−9484.
(206) Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987**, *4*, 406−425.
(207) Nelson, K. E.; Clayton, R. A.; Gill, S. R.; Gwinn, M. L.; Dodson, R. J.; Haft, D. H.; Hickey, E. K.; Peterson, J. D.; Nelson, W. C.; Ketchum, K. A.; McDonald, L.; Utterback, T. R.; Malek, J. A.; Linher, K. D.; Garrett, M. M.; Stewart, A. M.; Cotton, M. D.; Pratt, M. S.; Phillips, C. A.; Richardson, D.; Heidelberg, J.; Sutton, G. G.; Fleischmann, R. D.; Eisen, J. A.; Fraser, C. M. *Nature* **1999**, *399*, 323−329.
(208) Figge, R. M.; Schubert, M.; Brinkmann, H.; Cerff, R. *Mol. Biol. Evol.* **1999**, *16*, 429−440.
(209) Nelson, K. E.; Paulsen, I. T.; Heidelberg, J. F.; Fraser, C. M. *Nat. Biotechnol.* **2000**, *18*, 1049−1054.
(210) Royo, J.; Gomez, E.; Hueros, G. *J. Biol. Chem.* **2000**, *275*, 24993−24999.
(211) Royo, J.; Gimez, E.; Hueros, G. *Trends Genet.* **2000**, *16*, 432−433.
(212) Linton, D.; Karlyshev, A. V.; Hitchen, P. G.; Morris, H. R.; Dell, A.; Gregson, N. A.; Wren, B. W. *Mol. Microbiol.* **2000**, *35*, 1120−1134.
(213) Kanamori, A.; Nakayama, J.; Fukuda, M. N.; Stallcup, W. B.; Sasaki, K.; Fukuda, M.; Hirabayashi, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2897−2902.
(214) Shi, W. X.; Chammas, R.; Varki, A. *Glycobiology* **1998**, *8*, 199−205.
(215) Cameron, R. A.; Mahairas, G.; Rast, J. P.; Martinez, P.; Biondi, T. R.; Swartzell, S.; Wallace, J. C.; Poustka, A. J.; Livingston, B. T.; Wray, G. A.; Ettensohn, C. A.; Lehrach, H.; Britten, R. J.; Davidson, E. H.; Hood, L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9514−9518.
(216) Kawano, T.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A. *J. Biol. Chem.* **1994**, *269*, 9024−9029.
(217) Kawano, T.; Kozutsumi, Y.; Takematsu, H.; Kawasaki, T.; Suzuki, A. *Glycoconjugate J.* **1993**, *10*, 109−115.
(218) Takematsu, H.; Kawano, T.; Koyama, S.; Kozutsumi, Y.; Suzuki, A.; Kawasaki, T. *J. Biochem. (Tokyo)* **1994**, *115*, 381−386.
(219) Malykh, Y. N.; Shaw, L.; Schauer, R. *Glycoconjugate J.* **1998**, *15*, 885−893.
(220) Irie, A.; Koyama, S.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A. *J. Biol. Chem.* **1998**, *273*, 15866−15871.
(221) Schlenzka, W.; Shaw, L.; Schauer, R. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1993**, *1161*, 131−138.
(222) Schlenzka, W.; Shaw, L.; Kelm, S.; Schmidt, C. L.; Bill, E.; Trautwein, A. X.; Lottspeich, F.; Schauer, R. *FEBS Lett.* **1996**, *385*, 197−200.
(223) Schmidt, C. L.; Shaw, L. *J. Bioenerg. Biomembr.* **2001**, *33*, 9−26.
(224) Diaz, S.; Higa, H. H.; Hayes, B. K.; Varki, A. *J. Biol. Chem.* **1989**, *264*, 19416−19426.
(225) Butor, C.; Diaz, S.; Varki, A. *J. Biol. Chem.* **1993**, *268*, 10197−10206.
(226) Shi, W. X.; Chammas, R.; Varki, A. *J. Biol. Chem.* **1996**, *271*, 15130−15138.
(227) Higa, H. H.; Butor, C.; Diaz, S.; Varki, A. *J. Biol. Chem.* **1989**, *264*, 19427−19434.
(228) Varki, A.; Diaz, S. *Anal. Biochem.* **1984**, *137*, 236−247.
(229) Kamerling, J. P.; Schauer, R.; Shukla, A. K.; Stoll, S.; van Halbeek, H.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1987**, *162*, 601−607.
(230) Vandamme-Feldhaus, V.; Schauer, R. *J. Biochem. (Tokyo)* **1998**, *124*, 111−121.
(231) Higa, H.; Varki, A. *J. Biol. Chem.* **1988**, *263*, 8872−8878.
(232) Kamerling, J. P.; Dorland, L.; van Halbeek, H.; Vliegenthart, J. F. G.; Messer, M.; Schauer, R. *Carbohydr. Res.* **1982**, *100*, 331−340.
(233) Corfield, A. P.; Ferreira do, Amaral. C.; Wember, M.; Schauer, R. *Eur. J. Biochem.* **1976**, *68*, 597−610.
(234) Iwersen, M.; Vandamme-Feldhaus, V.; Schauer, R. *Glycoconjugate J.* **1998**, *15*, 895−904.
(235) Higashi, H.; Hirabayashi, Y.; Fukui, Y.; Naiki, M.; Matsumoto, M.; Ueda, S.; Kato, S. *Cancer Res.* **1985**, *45*, 3796−3802.
(236) Kelm, A.; Shaw, L.; Schauer, R.; Reuter, G. *Eur. J. Biochem.* **1998**, *251*, 874−884.
(237) Guimaraes, M. J.; Bazan, J. F.; Castagnola, J.; Diaz, S.; Copeland, N. G.; Gilbert, D. J.; Jenkins, N. A.; Varki, A.; Zlotnik, A. *J. Biol. Chem.* **1996**, *271*, 13697−13705.
(238) Takematsu, H.; Diaz, S.; Stoddart, A.; Zhang, Y.; Varki, A. *J. Biol. Chem.* **1999**, *274*, 25623−25631.
(239) Milner, C. M.; Smith, S. V.; Carrillo, M. B.; Taylor, G. L.; Hollinshead, M.; Campbell, R. D. *J. Biol. Chem.* **1997**, *272*, 4549−4558.
(240) Pshezhetsky, A. V.; Richard, C.; Michaud, L.; Igdoura, S.; Wang, S. P.; Elsliger, M. A.; Qu, J. Y.; Leclerc, D.; Gravel, R.; Dallaire, L.; Potier, M. *Nat. Genet.* **1997**, *15*, 316−320.
(241) Hasegawa, T.; Yamaguchi, K.; Wada, T.; Takeda, A.; Itoyama, Y.; Miyagi, T. *J. Biol. Chem.* **2000**, *275*, 8007−8015.
(242) Miyagi, T.; Konno, K.; Emori, Y.; Kawasaki, H.; Suzuki, K.; Yasui, A.; Tsuiki, S. *J. Biol. Chem.* **1993**, *268*, 26435−26440.
(243) Verheijen, F. W.; Verbeek, E.; Aula, N.; Beerens, C. E. M. T.; Havelaar, A. C.; Joosse, M.; Peltonen, L.; Aula, P.; Galjaard, H.; Van der Spek, P. J.; Mancini, G. M. S. *Nat. Genet.* **1999**, *23*, 462−465.
(244) Schauer, R.; Wember, M. *Biol. Chem. Hoppe Seyler* **1996**, *377*, 293−299.
(245) de Koning, A. P.; Brinkman, F. S.; Jones, S. J.; Keeling, P. J. *Mol. Biol. Evol.* **2000**, *17*, 1769−1773.
(246) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Funke, R.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Gage, D.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Harris, K.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Heaford, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Howland, J.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Kann, L.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Lehoczky, J.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; LeVine, R.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; McEwan, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; McKernan, K.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Meldrim, J.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Mesirov, J. P.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Miranda, C.; Smit, A. F.; Stupka, E.; Szustakowki, J.; Thierry-Mieg, D.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Morris, W.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E. *Nature* **2001**, *409*, 860−921.
(247) Salzberg, S. L.; White, O.; Peterson, J.; Eisen, J. A. *Science* **2001**, *292*, 1903−1906.
(248) Doolittle, W. F. *Trends Genet.* **1998**, *14*, 307−311.
(249) Bibb, M. *Microbiology* **1996**, *142*, 1335−1344.
(250) Shintani, S.; Terzic, J.; Sato, A.; Saraga-Babic, M.; O'hUigin, C.; Tichy, H.; Klein, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7417−7422.
(251) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(252) Rogers, G. N.; Paulson, J. C. *Virology* **1983**, *127*, 361−373.

(253) Ito, T.; Suzuki, Y.; Mitnaul, L.; Vines, A.; Kida, H.; Kawaoka, Y. *Virology* **1997**, *227*, 493−499.
(254) Xu, G.; Suzuki, T.; Tahara, H.; Kiso, M.; Hasegawa, A.; Suzuki, Y. *J. Biochem. (Tokyo)* **1994**, *115*, 202−207.
(255) Harms, G.; Reuter, G.; Corfield, A. P.; Schauer, R. *Glycoconjugate J.* **1996**, *13*, 621−630.
(256) Suzuki, Y.; Suzuki, T.; Matsunaga, M.; Matsumoto, M. *J. Biochem. (Tokyo)* **1985**, *97*, 1189−1199.
(257) Paulson, J. C.; Sadler, J. E.; Hill, R. L. *J. Biol. Chem.* **1979**, *254*, 2120−2124.
(258) Markwell, M. A.; Svennerholm, L.; Paulson, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 5406−5410.
(259) Reyes-Leyva, J.; Espinosa, B.; Santos, G.; Zenteno, R.; Hernández, J.; Vallejo, V.; Zenteno, E. *Glycoconjugate J.* **1999**, *16*, 517−522.
(260) Barton, E. S.; Connolly, J. L.; Forrest, J. C.; Chappell, J. D.; Dermody, T. S. *J. Biol. Chem.* **2001**, *276*, 2200−2211.
(261) Gentsch, J. R.; Pacitti, A. F. *J. Virol.* **1985**, *56*, 356−364.
(262) Rolsma, M. D.; Kuhlenschmidt, T. B.; Gelberg, H. B.; Kuhlenschmidt, M. S. *J. Virol.* **1998**, *72*, 9079−9091.
(263) Svensson, L. *J. Virol.* **1992**, *66*, 5582−5585.
(264) Guo, C. T.; Nakagomi, O.; Mochizuki, M.; Ishida, H.; Kiso, M.; Ohta, Y.; Suzuki, T.; Miyamoto, D.; Hidari, K. I. P. J.; Suzuki, Y. *J. Biochem. (Tokyo)* **1999**, *126*, 683−688.
(265) Kuhlenschmidt, T. B.; Hanafin, W. P.; Gelberg, H. B.; Kuhlenschmidt, M. S. *Adv. Exp. Med. Biol.* **1999**, *473*, 309−317.
(266) Ludert, J. E.; Mason, B. B.; Angel, J.; Tang, B. Z.; Hoshino, Y.; Feng, N. G.; Vo, P. T.; Mackow, E. M.; Ruggeri, F. M.; Greenberg, H. B. *J. Gen. Virol.* **1998**, *79*, 725−729.
(267) Delorme, C.; Brüssow, H.; Sidoti, J.; Roche, N.; Karlsson, K. A.; Neeser, J. R.; Teneberg, S. *J. Virol.* **2001**, *75*, 2276−2287.
(268) Eaton, B. T.; Crameri, G. S. *J. Gen. Virol.* **1989**, *70*, 3347−3353.
(269) Vlasak, R.; Luytjes, W.; Spaan, W.; Palese, P. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4526−4529.
(270) Kunkel, F.; Herrler, G. *Virology* **1993**, *195*, 195−202.
(271) Schultze, B.; Gross, H.-J.; Brossmer, R.; Klenk, H.-D.; Herrler, G. *Virus Res.* **1990**, *16*, 185−194.
(272) Schultze, B.; Krempl, C.; Ballesteros, M. L.; Shaw, L.; Schauer, R.; Enjuanes, L.; Herrler, G. *J. Virol.* **1996**, *70*, 5634−5637.
(273) Schultze, B.; Cavanagh, D.; Herrler, G. *Virology* **1992**, *189*, 792−794.
(274) Arnberg, N.; Edlund, K.; Kidd, A. H.; Wadell, G. *J. Virol.* **2000**, *74*, 42−48.
(275) Arnberg, N.; Kidd, A. H.; Edlund, K.; Olfat, F.; Wadell, G. *J. Virol.* **2000**, *74*, 7691−7693.
(276) Stoner, G. D.; Williams, B.; Kniazeff, A.; Shimkin, M. B. *Nature* **1973**, *245*, 319−320.
(277) Utagawa, E. T.; Miyamura, K.; Mukoyama, A.; Kono, R. *J. Gen. Virol.* **1982**, *63*, 141−148.
(278) Zhou, L.; Luo, Y.; Wu, Y.; Tsao, J.; Luo, M. *J. Virol.* **2000**, *74*, 1477−1485.
(279) Uncapher, C. R.; DeWitt, C. M.; Colonno, R. J. *Virology* **1991**, *180*, 814−817.
(280) Tavakkol, A.; Burness, A. T. H. *Biochemistry* **1990**, *29*, 10684−10690.
(281) Enegren, B. J.; Burness, A. T. *Nature.* **1977**, *268*, 536−537.
(282) Anderson, K.; Bond, C. W. *Arch. Virol.* **1987**, *93*, 31−49.
(283) Kaludov, N.; Brown, K. E.; Walters, R. W.; Zabner, J.; Chiorini, J. A. *J. Virol.* **2001**, *75*, 6884−6893.
(284) Walters, R. W.; Yi, S. M. P.; Keshavjee, S.; Brown, K. E.; Welsh, M. J.; Chiorini, J. A.; Zabner, J. *J. Biol. Chem.* **2001**, *276*, 20610−20616.
(285) Barbis, D. P.; Chang, S.-F.; Parrish, C. R. *Virology* **1992**, *191*, 301−308.
(286) Thacker, T. C.; Johnson, F. B. *J. Gen. Virol.* **1998**, *79*, 2163-2169.
(287) Cahan, L. D.; Singh, R.; Paulson, J. C. *Virology* **1983**, *130*, 281−289.
(288) Fried, H.; Cahan, L. D.; Paulson, J. C. *Virology* **1981**, *109*, 188−192.
(289) Liu, C. K.; Wei, G.; Atwood, W. J. *J. Virol.* **1998**, *72*, 4643−4649.
(290) Sinibaldi, L.; Goldoni, P.; Pietropaolo, V.; Longhi, C.; Orsi, N. *Arch. Virol.* **1990**, *113*, 291−296.
(291) Keppler, O. T.; Herrmann, M.; Oppenländer, M.; Meschede, W.; Pawlita, M. *J. Virol.* **1994**, *68*, 6933−6939.
(292) Keppler, O. T.; Stehling, P.; Herrmann, M.; Kayser, H.; Grunow, D.; Reutter, W.; Pawlita, M. *J. Biol. Chem.* **1995**, *270*, 1308−1314.
(293) Lobert, P. E.; Hober, D.; Dewilde, A.; Wattré, P. *Arch. Virol.* **1995**, *140*, 1357−1371.
(294) Ravindranath, R. M. H.; Graves, M. C. *J. Virol.* **1990**, *64*, 5430−5440.
(295) Superti, F.; Hauttecoeur, B.; Morelec, M. J.; Goldoni, P.; Bizzini, B.; Tsiang, H. *J. Gen. Virol.* **1986**, *67*, 47−56.
(296) Conti, C.; Mastromarino, P.; Ciuffarella, M. G.; Orsi, N. *Arch. Virol.* **1988**, *99*, 261−269.
(297) Komai, K.; Kaplan, M.; Peeples, M. E. *Virology* **1988**, *163*, 629−634.
(298) Hacker, J.; Kestler, H.; Hoschutzky, H.; Jann, K.; Lottspeich, F.; Korhonen, T. K. *Infect. Immun.* **1993**, *61*, 544−550.
(299) Hanisch, F. G.; Hacker, J.; Schroten, H. *Infect. Immun.* **1993**, *61*, 2108−2115.
(300) Brooks, D. E.; Cavanagh, J.; Jayroe, D.; Janzen, J.; Snoek, R.; Trust, T. J. *Infect. Immun.* **1989**, *57*, 377−383.
(301) Lindahl, M.; Faris, A.; Wadstrom, T. *Lancet* **1982**, *2* (8292), 280.
(302) Bartus, H.; Actor, P.; Snipes, E.; Sedlock, D.; Zajac, I. *J. Clin. Microbiol.* **1985**, *21*, 951−954.
(303) Sjöberg, P.-O.; Lindahl, M.; Porath, J.; Wadström, T. *Biochem. J.* **1988**, *255*, 105−111.
(304) Miller-Podraza, H.; Abul Milh, M.; Teneberg, S.; Karlsson, K. A. *Infect. Immun.* **1997**, *65*, 2480−2482.
(305) Karlsson, K. A. *Glycobiology* **2000**, *10*, 761−771.
(306) Nyberg, G.; Strömberg, N.; Jonsson, A.; Karlsson, K.-A.; Normark, S. *Infect. Immun.* **1990**, *58*, 2555−2563.
(307) Scharfman, A.; Degroote, S.; Beau, J.; Lamblin, G.; Roussel, P.; Mazurier, J. *Glycobiology* **1999**, *9*, 757−764.
(308) Hazlett, L.; Rudner, X.; Masinick, S.; Ireland, M.; Gupta, S. *Invest. Ophthalmol. Vis. Sci.* **1995**, *36*, 634−643.
(309) Kawakami, K.; Ahmed, K.; Utsunomiya, Y.; Rikitomi, N.; Hori, A.; Oishi, K.; Nagatake, T. *Microbiol. Immunol.* **1998**, *42*, 697−702.
(310) Fakih, M. G.; Murphy, T. F.; Pattoli, M. A.; Berenson, C. S. *Infect. Immun.* **1997**, *65*, 1695−1700.
(311) St Geme, J. W. *Infect. Immun.* **1994**, *62*, 3881−3889.
(312) Ishikawa, H.; Isayama, Y. *Infect. Immun.* **1987**, *55*, 1607−1609.
(313) Arp, L. H.; Huffman, E. L.; Hellwig, D. H. *Am. J. Vet. Res.* **1993**, *54*, 2027−2030.
(314) Ahmed, K.; Matsumoto, K.; Rikitomi, N.; Nagatake, T. *FEMS Microbiol. Lett.* **1996**, *135*, 305−309.
(315) Groenink, J.; Ligtenberg, A. J.; Veerman, E. C.; Bolscher, J. G.; Nieuw Amerongen, A. V. *Antonie Van Leeuwenhoek* **1996**, *70*, 79−87.
(316) Neeser, J.-R.; Grafström, R. C.; Woltz, A.; Brassart, D.; Fryder, V.; Guggenheim, B. *Glycobiology* **1995**, *5*, 97−104.
(317) Murray, P. A.; Levine, M. J.; Tabak, L. A.; Reddy, M. S. *Biochem. Biophys. Res. Commun.* **1982**, *106*, 390−396.
(318) Takahashi, Y.; Sandberg, A. L.; Ruhl, S.; Muller, J.; Cisar, J. O. *Infect. Immun.* **1997**, *65*, 5042−5051.
(319) Oho, T.; Yu, H.; Yamashita, Y.; Koga, T. *Infect. Immun.* **1998**, *66*, 115−121.
(320) Liukkonen, J.; Haataja, S.; Tikkanen, K.; Kelm, S.; Finne, J. *J. Biol. Chem.* **1992**, *267*, 21105−21111.
(321) Rosenow, C.; Ryan, P.; Weiser, J. N.; Johnson, S.; Fontan, P.; Ortqvist, A.; Masure, H. R. *Mol. Microbiol.* **1997**, *25*, 819−829.
(322) Thomas, V. L.; Sanford, B. A.; Ramsay, M. A. *J. Gen. Microbiol.* **1993**, *139*, 623−629.
(323) Smith, D. G.; Russell, W. C.; Thirkell, D. *Microbiology* **1994**, *140*, 2893−2898.
(324) Domingues, R. M.; Cavalcanti, S. M.; Andrade, A. F.; Ferreira, M. C. *Int. J. Med. Microbiol. Virol. Parasitol. Infect. Dis.* **1992**, *277*, 340−344.
(325) Mikx, F. H.; Keulers, R. A. *Infect. Immun.* **1992**, *60*, 1761−1766.
(326) Steiner, B. M.; Sell, S.; Schell, R. F. *J. Infect. Dis.* **1987**, *155*, 742−748.
(327) Loomes, L. M.; Uemura, K.; Childs, R. A.; Paulson, J. C.; Rogers, G. N.; Scudder, P. R.; Michalski, J. C.; Hounsell, E. F.; Taylor Robinson, D.; Feizi, T. *Nature* **1984**, *307*, 560−563.
(328) Sachse, K.; Pfutzner, H.; Heller, M.; Hanel, I. *Vet. Microbiol.* **1993**, *36*, 307−316.
(329) Glasgow, L. R.; Hill, R. L. *Infect. Immun.* **1980**, *30*, 353−361.
(330) MacKenzie, C. R.; Hirama, T.; Lee, K. K.; Altman, E.; Young, N. M. *J. Biol. Chem.* **1997**, *272*, 5533−5538.
(331) Takamizawa, K.; Iwamori, M.; Kozaki, S.; Sakaguchi, G.; Tanaka, R.; Takayama, H.; Nagai, Y. *FEBS Lett.* **1986**, *201*, 229−232.
(332) Kozaki, S.; Kamata, Y.; Watarai, S.; Nishiki, T.; Mochida, S. *Microb. Pathog.* **1998**, *25*, 91−99.
(333) Wadsworth, J. D.; Desai, M.; Tranter, H. S.; King, H. J.; Hambleton, P.; Melling, J.; Dolly, J. O.; Shone, C. C. *Biochem. J.* **1990**, *268*, 123−128.
(334) Jolivet-Reynaud, C.; Hauttecoeur, B.; Alouf, J. E. *Toxicon* **1989**, *27*, 1113−1126.
(335) Nagahama, M.; Sakurai, J. *Infect. Immun.* **1992**, *60*, 1237−1240.
(336) Holmgren, J.; Lonnroth, I.; Mansson, J.; Svennerholm, L. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2520−2524.
(337) Shinoda, S.; Ishida, K.; Oh, E. G.; Sasahara, K.; Miyoshi, S.; Chowdhury, M. A.; Yasuda, T. *Microbiol. Immunol.* **1993**, *37*, 405−409.
(338) Takeda, Y.; Takeda, T.; Honda, T.; Miwatani, T. *Infect. Immun.* **1976**, *14*, 1−5.
(339) Hausman, S. Z.; Burns, D. L. *Infect. Immun.* **1993**, *61*, 335−337.
(340) Armstrong, G. D.; Howard, L. A.; Peppler, M. S. *J. Biol. Chem.* **1988**, *263*, 8677−8684.
(341) Kato, I.; Naiki, M. *Infect. Immun.* **1976**, *13*, 289−291.
(342) Nariya, H.; Kamio, Y. *Biosci., Biotechnol., Biochem.* **1995**, *59*, 1603−1604.

(343) Babál, P.; Pindak, F. F.; Wells, D. J.; Gardner, W. A., Jr. *Biochem. J.* **1994**, *299*, 341−346.
(344) Pakandl, M.; Grubhoffer, L. *Comp. Biochem. Physiol. [B].* **1994**, *108*, 529−536.
(345) Feingold, C.; Bracha, R.; Wexler, A.; Mirelman, D. *Infect. Immun.* **1985**, *48*, 211−218.
(346) Monteiro, V. G.; Soares, C. P.; De Souza, W. *FEMS Microbiol. Lett.* **1998**, *164*, 323−327.
(347) Bouchara, J. P.; Sanchez, M.; Chevailler, A.; Marot-Leblond, A.; Lissitzky, J. C.; Tronchin, G.; Chabasse, D. *Infect. Immun.* **1997**, *65*, 2717−2724.
(348) Hamilton, A. J.; Jeavons, L.; Youngchim, S.; Vanittanakom, N. *Infect. Immun.* **1999**, *67*, 5200−5205.
(349) Chabasse, D.; Robert, R. *Ann. Inst. Pasteur Microbiol.* **1986**, *137B*, 187−193.
(350) Bouchara, J. P.; Robert, R.; Chabasse, D.; Senet, J. M. *Ann. Inst. Pasteur Microbiol.* **1987**, *138*, 729−736.
(351) Varki, A. *J. Clin. Invest.* **1997**, *99*, 158−162.
(352) Crocker, P. R.; Mucklow, S.; Bouckson, V.; McWilliam, A.; Willis, A. C.; Gordon, S.; Milon, G.; Kelm, S.; Bradfield, P. *EMBO J.* **1994**, *13*, 4490−4503.
(353) Powell, L. D.; Varki, A. *J. Biol. Chem.* **1994**, *269*, 10628−10636.
(354) Powell, L. D.; Jain, R. K.; Matta, K. L.; Sabesan, S.; Varki, A. *J. Biol. Chem.* **1995**, *270*, 7523−7532.
(355) Freeman, S. D.; Kelm, S.; Barber, E. K.; Crocker, P. R. *Blood* **1995**, *85*, 2005−2012.
(356) Collins, B. E.; Yang, L. J. S.; Mukhopadhyay, G.; Filbin, M. T.; Kiso, M.; Hasegawa, A.; Schnaar, R. L. *J. Biol. Chem.* **1997**, *272*, 1248−1255.
(357) Collins, B. E.; Ito, H.; Sawada, N.; Ishida, H.; Kiso, M.; Schnaar, R. L. *J. Biol. Chem.* **1999**, *274*, 27899−27893.
(358) Cornish, A. L.; Freeman, S.; Forbes, G.; Ni, J.; Zhang, M.; Cepeda, M.; Gentz, R.; Augustus, M.; Carter, K. C.; Crocker, P. R. *Blood* **1998**, *92*, 2123-2132.
(359) Patel, N.; Brinkman-Van der Linden, E. C. M.; Altmann, S. W.; Gish, K.; Balasubramanian, S.; Timans, J. C.; Peterson, D.; Bell, M. P.; Bazan, J. F.; Varki, A.; Kastelein, R. A. *J. Biol. Chem.* **1999**, *274*, 22729−22738.
(360) Nicoll, G.; Ni, J.; Liu, D.; Klenerman, P.; Munday, J.; Dubock, S.; Mattei, M. G.; Crocker, P. R. *J. Biol. Chem.* **1999**, *274*, 34089−34095.
(361) Angata, T.; Varki, A. *Glycobiology* **2000**, *10*, 431−438.
(362) Floyd, H.; Ni, J.; Cornish, A. L.; Zeng, Z. Z.; Liu, D.; Carter, K. C.; Steel, J.; Crocker, P. R. *J. Biol. Chem.* **2000**, *275*, 861−866.
(363) Zhang, J. Q.; Nicoll, G.; Jones, C.; Crocker, P. R. *J. Biol. Chem.* **2000**, *275*, 22121−22126.
(364) Munday, J.; Kerr, S.; Ni, J.; Cornish, A. L.; Zhang, J. Q.; Nicoll, G.; Floyd, H.; Mattei, M. G.; Moore, P.; Liu, D.; Crocker, P. R. *Biochem. J.* **2001**, *355*, 489−497.
(365) Fearon, D. T. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 1971−1975.
(366) Pangburn, M. K.; Muller-Eberhard, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2416−2420.
(367) Ravindranath, M. H.; Gonzales, A.; Soh, D.; Nishimoto, K.; Tam, W. Y.; Bilchik, A.; Morton, D. L.; O'Day, S. *Biochem. Biophys. Res. Commun.* **2001**, *283*, 369−373.
(368) Banerjee, M.; Chowdhury, M. *Hum. Reprod.* **1994**, *9*, 1497−1504.
(369) Chakraborty, I.; Mandal, C.; Chowdhury, M. *Mol. Cell. Biochem.* **1993**, *126*, 77−86.
(370) Nitta, K.; Takayanagi, G.; Kawauchi, H.; Hakomori, S. *Cancer Res.* **1987**, *47*, 4877−4883.
(371) Tiemeyer, M.; Yasuda, Y.; Schnaar, R. L. *J. Biol. Chem.* **1989**, *264*, 1671−1681.
(372) Tiemeyer, M.; Swank-Hill, P.; Schnaar, R. L. *J. Biol. Chem.* **1990**, *265*, 11990−11999.
(373) Popoli, M.; Mengano, A. *Neurochem. Res.* **1988**, *13*, 63−67.
(374) Miller, R. L.; Collawn, J. F. Jr.; Fish, W. W. *J. Biol. Chem.* **1982**, *257*, 7574−7580.
(375) Sen, G.; Mandal, C. *Carbohydr. Res.* **1995**, *268*, 115−125.
(376) Brossmer, R.; Wagner, M.; Fischer, E. *J. Biol. Chem.* **1992**, *267*, 8752−8756.
(377) Swarnakar, S.; Sarathi Chowdhury, P.; Sarkar, M. *Biochem. Biophys. Res. Commun.* **1991**, *178*, 85−94.
(378) Bretting, H.; Stanislawski, E.; Jacobs, G.; Becker, W. *Biochim. Biophys. Acta* **1983**, *749*, 143−152.
(379) Dam, T. K.; Sarkar, M.; Ghosal, J.; Choudhury, A. *Biochem. Biophys. Res. Commun.* **1993**, *196*, 422−429.
(380) Tunkijjanukij, S.; Mikkelsen, H. V.; Olafsen, J. A. *Comp. Biochem. Physiol. [B]* **1997**, *117B*, 273−286.
(381) Hardy, S. W.; Grant, P. T.; Fletcher, T. C. *Experientia.* **1977**, *33*, 767−769.
(382) Hardy, S. W.; Fletcher, T. C.; Gerrie, L. M. *Biochem. Soc. Trans.* **1976**, *4*, 473−475.
(383) Muresan, V.; Iwanij, V.; Smith, Z. D.; Jamieson, J. D. *J. Histochem. Cytochem.* **1982**, *30*, 938−946.
(384) Mohan, S.; Thambi Dorai, D.; Srimal, S.; Bachhawat, B. K. *Biochem. J.* **1982**, *203*, 253−261.
(385) Dorai, D. T.; Mohan, S.; Srimal, S.; Bachhawat, B. K. *FEBS Lett.* **1982**, *148*, 98−102.
(386) Tsuboi, I.; Matsukawa, M.; Sato, N.; Kimura, S. *Biochim. Biophys. Acta Gen. Subj.* **1993**, *1156*, 255−262.
(387) Iwaki, D.; Osaki, T.; Mizunoe, Y.; Wai, S. N.; Iwanaga, S.; Kawabata, S. *Eur. J. Biochem.* **1999**, *264*, 314−326.
(388) Fischer, E.; Quoc Khang, N.; Letendre, G.; Brossmer, R. *Glycoconjugate J.* **1994**, *11*, 51−58.
(389) Tsuboi, I.; Matsukawa, M.; Sato, N. *Biosci. Biotechnol. Biochem.* **1993**, *57*, 1237−1242.
(390) Vasta, G. R.; Ilodi, G. H.; Cohen, E.; Brahmi, Z. *Dev. Comp. Immunol.* **1982**, *6*, 625−634.
(391) Vasta, G. R.; Cohen, E. *Dev. Comp. Immunol.* **1982**, *6*, 219−230.
(392) Vasta, G. R.; Cohen, E. *J. Invertebr. Pathol.* **1984**, *43*, 333−342.
(393) Ahmed, H.; Chatterjee, B. P.; Kelm, S.; Schauer, R. *Biol. Chem. Hoppe Seyler* **1986**, *367*, 501−506.
(394) Vasta, G. R.; Cohen, E. *Experientia* **1983**, *39*, 721−722.
(395) Vasta, G. R.; Cohen, E. *Experientia* **1984**, *40*, 485−487.
(396) Vasta, G. R.; Cohen, E. *Comp. Biochem. Physiol.* **1984**, *77B*, 721−727.
(397) Vasta, G. R.; Cohen, E. *Dev. Comp. Immunol.* **1984**, *8*, 515−522.
(398) Kuhn, K. H.; Uhlir, J.; Grubhoffer, L. *Parasitol. Res.* **1996**, *82*, 215−221.
(399) Kovar, V.; Kopacek, P.; Grubhoffer, L. *Insect Biochem. Mol. Biol.* **2000**, *30*, 195−205.
(400) Ravindranath, M. H.; Higa, H. H.; Cooper, E. L.; Paulson, J. C. *J. Biol. Chem.* **1985**, *260*, 8850−8856.
(401) Vasta, G. R.; Cohen, E. *Dev. Comp. Immunol.* **1984**, *8*, 197−202.
(402) Mercy, S. P. D.; Ravindranath, M. H. *Eur. J. Biochem.* **1993**, *215*, 697−704.
(403) Abel, C. A.; Campbell, P. A.; VanderWall, J.; Hartman, A. L. *Prog. Clin. Biol. Res.* **1984**, *157*, 103−114.
(404) Kopacek, P.; Grubhoffer, L.; Soderhall, K. *Dev. Comp. Immunol.* **1993**, *17*, 407−418.
(405) Ratanapo, S.; Chulavatnatol, M. *Comp. Biochem. Physiol. [B]* **1990**, *97B*, 515−520.
(406) Zenteno, R.; Vazquez, L.; Martinez-Cairo, S.; Bouquelet, S.; Agundis, C.; Zenteno, E. *Glycoconjugate J.* **2000**, *17*, 339−347.
(407) Yamashita, K.; Umetsu, K.; Suzuki, T.; Iwaki, Y.; Endo, T.; Kobata, A. *J. Biol. Chem.* **1988**, *263*, 17482−17489.
(408) Hapner, K. D.; Jermyn, M. A. *Insect Biochem.* **1981**, *11*, 287−296.
(409) Bettencourt, R.; Gunne, H.; Gastinel, L.; Steiner, H.; Faye, I. *Eur. J. Biochem.* **1999**, *266*, 964−976.
(410) Vasta, G. R.; Warr, G. W.; Marchalonis, J. J. *Comp. Biochem. Physiol.* **1982**, *73B*, 887−900.
(411) Shibuya, N.; Goldstein, I. J.; Broekaert, W. F.; Nsimba-Lubaki, M.; Peeters, B.; Peumans, W. J. *J. Biol. Chem.* **1987**, *262*, 1596−1601.
(412) Van Damme, E. J. M.; Roy, S.; Barre, A.; Citores, L.; Mostafapous, K.; Rougé, P.; Van Leuven, F.; Girbés, T.; Goldstein, I. J.; Peumans, W. J. *Eur. J. Biochem.* **1997**, *245*, 648−655.
(413) Shibuya, N.; Tazaki, K.; Song, Z.; Tarr, G. E.; Goldstein, I. J.; Peumans, W. J. *J. Biochem. (Tokyo)* **1989**, *106*, 1098−1103.
(414) Kaku, H.; Tanaka, Y.; Tazaki, K.; Minami, E.; Mizuno, H.; Shibuya, N. *J. Biol. Chem.* **1996**, *271*, 1480−1485.
(415) Knibbs, R. N.; Goldstein, I. J.; Ratcliffe, R. M.; Shibuya, N. *J. Biol. Chem.* **1991**, *266*, 83−88.
(416) Imberty, A.; Gautier, C.; Lescar, J.; Pérez, S.; Wyns, L. *J. Biol. Chem.* **2000**, *275*, 17541−17548.
(417) Konami, Y.; Yamamoto, K.; Osawa, T.; Irimura, T. *FEBS Lett.* **1994**, *342*, 334−338.
(418) Yamashita, K.; Umetsu, K.; Suzuki, T.; Ohkura, T. *Biochemistry* **1992**, *31*, 11647−11650.
(419) Ray, S.; Chatterjee, B. P. *Phytochemistry* **1995**, *40*, 643−649.
(420) Ratanapo, S.; Wayakorn, N.; Ngamjunyaporn, W.; Chulavatnatol, M. *Plant Sci.* **1998**, *139*, 141−148.
(421) Yen, S. E.; Mansfield, J. M.; Wallace, J. H. *Int. Arch. Allergy Appl. Immunol.* **1980**, *61*, 32−39.
(422) Debray, H.; Montreuil, J.; Franz, H. *Glycoconjugate J.* **1994**, *11*, 550−557.
(423) Wu, A. M.; Song, S. C.; Hwang, P. Y.; Wu, J. H.; Pfüller, U. *Biochem. Biophys. Res. Commun.* **1995**, *214*, 396−402.
(424) Monsigny, M.; Roche, A. C.; Sene, C.; Maget Dana, R.; Delmotte, F. *Eur. J. Biochem.* **1980**, *104*, 147−153.
(425) Kawagishi, H.; Mori, H.; Uno, A.; Kimura, A.; Chiba, S. *FEBS Lett.* **1994**, *340*, 56−58.
(426) Yagi, F.; Miyamoto, M.; Abe, T.; Minami, Y.; Tadera, K.; Goldstein, I. J. *Glycoconjugate J.* **1997**, *14*, 281−288.
(427) Mo, H. Q.; Winter, H. C.; Goldstein, I. J. *J. Biol. Chem.* **2000**, *275*, 10623−10629.
(428) Baldauf, S. L.; Palmer, J. D.; Doolittle, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7749−7754.

CR000407M