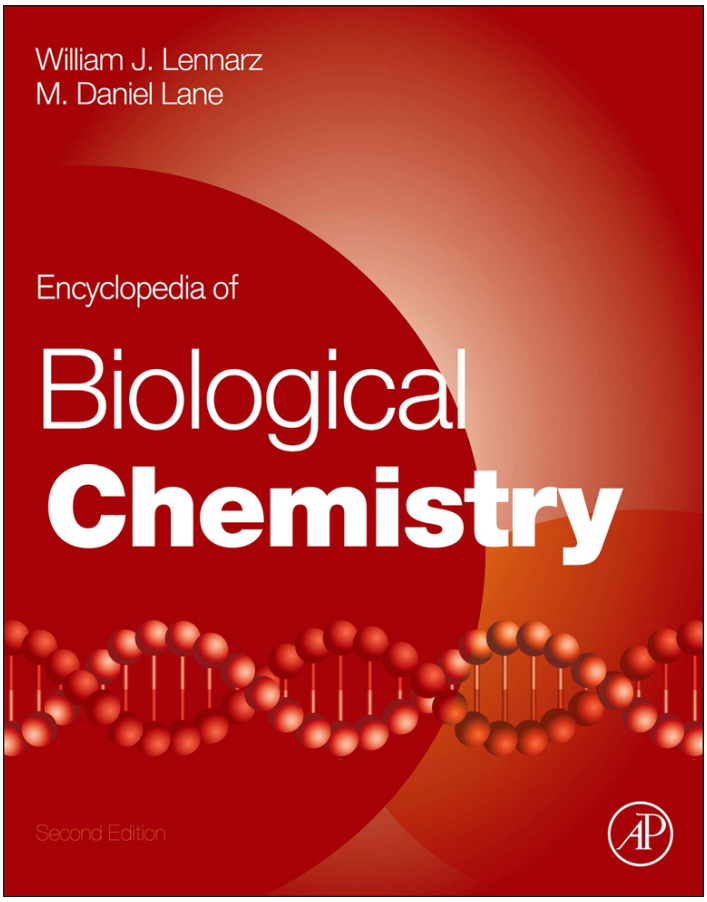

Varki A. (2013) Siglecs. In: Lennarz W.J. and Lane M.D. (eds.) The Encyclopedia of Biological Chemistry, vol. 4, pp. 229-231. Waltham, MA: Academic Press.

Siglecs

A Varki, University of California, San Diego, CA, USA

Glossary

Immunoglobulin superfamily (IgSf) Proteins that have modules homologous to those of antibodies (immunoglobulins). This is an evolutionarily ancient group of proteins whose appearance actually predated the emergence of the immunoglobulins themselves.

I-type lectins Proteins (other than antibodies or T-cell receptors) in which immunoglobulin-like modules mediate binding to glycans (sugar chains).

Sialic acids These acids are a diverse family of nine-carbon acidic sugars that typically occupy a terminal position on glycan chains attached to the cell surface of higher animals of the deuterostome lineage.

Siglecs A major subset of the I-type lectins. Name is based on their defining properties, as sialic acid recognizing IgSf lectins with canonical amino-terminal sequences.

Historical Background and Definition

Sialic acids (Sias) are a family of nine-carbon acidic sugars that typically occupy a terminal position on glycan chains attached to the cell surface of higher animals. The immunoglobulin superfamily (IgSf) is an evolutionarily ancient group of proteins whose appearance predated the emergence of the immunoglobulins themselves. Until the 1990s, it was assumed that IgSf members (other than some antibodies) did not mediate carbohydrate recognition. Independent work on sialoadhesin (Sn, eventually Siglec-1, a protein on certain macrophage subsets) and on CD22 (eventually Siglec-2, a protein on mature B cells) revealed that their first Ig V-set-like domains could mediate Sia recognition. Homologous features of this V-set Ig-like domain and the adjacent C2-set domain then led to the discovery that two other previously cloned molecules – CD33 (eventually Siglec-3) and myelin-associated glycoprotein (MAG, eventually Siglec-4) – also had Sia-binding properties. Following consultation among all researchers working on these proteins, the common name Siglec (suggested by A. Varki) and a numbering system were agreed upon. Criteria for inclusion of IgSf-related proteins as Siglecs are: (1) the ability to recognize sialylated glycans and (2) significant sequence similarity within the N-terminal V-set and adjoining C2-set domains. Evaluation of the human and mouse genomes eventually defined 16 primate and eight mouse molecules that fulfill these criteria. Primates have many more Siglecs than rodents and many of these are rapidly evolving.

Two Broad Subgroups of Siglecs

While Siglecs -1, -2, -4, and -15 appear to be evolutionarily rather conserved, the CD33/Siglec-3-related subgroup (Primate Siglecs -3 and -5–16) are rapidly evolving. Some CD33/Siglec-3-related Siglecs appear to have evolved as hybrids of pre-existing genes and/or by gene conversion. For these reasons, sequence comparisons alone do not allow the conclusive designation of the ortholog status of all genes, and additional features such as gene position and exon structure must be taken into account. As such issues cannot always be resolved, some mouse Siglecs have been assigned an alphabetical designation.

Common Structural Features

All Siglecs are single-pass type 1 integral membrane proteins with extracellular domains consisting of uniquely similar N-terminal V-set Ig domains, followed by variable numbers of C2-set Ig domains, ranging from 16 in Sn/Siglec-1 to 1 in CD33/Siglec-3. Crystal structures for mouse Siglec-1 and human Siglec-7 indicate (Figure 1) that the V-set immunoglobulin-like fold has several unusual features, including an intra-beta sheet disulfide and a splitting of the standard beta strand G into two shorter strands. These features along with certain key amino acid residues appear to be requirements for Sia recognition. In particular, a conserved arginine residue is involved in a salt bridge with the carboxylate of Sia, in all instances studied to date.

Cell-Type Specific Expression

With the exception of MAG/Siglec-4, Siglec-6, and Siglec-12, expression so far appears to be confined to the hematopoeitic and immune systems. Within these systems, each Siglec is expressed in a cell-type specific fashion, suggesting that each may be involved in discrete functions. However, systematic studies of Siglec expression outside the hematopoeitic system and during development have not yet been done.

Genomic Organization and Phylogeny

Based on searching for the canonical functional amino acids in the V-set domain of the typical Siglec, there is so far no evidence for Siglec-like molecules in prokaryotes, fungi, or plants, or in animals of the protostome lineage, including organisms for which the complete genome is available. By contrast, it is relatively easy to find Siglec-like V-set domains in many vertebrate taxa. While the conserved Siglecs (-1, -2, -4, and -15) have relatively clear-cut single orthologs that can be identified in

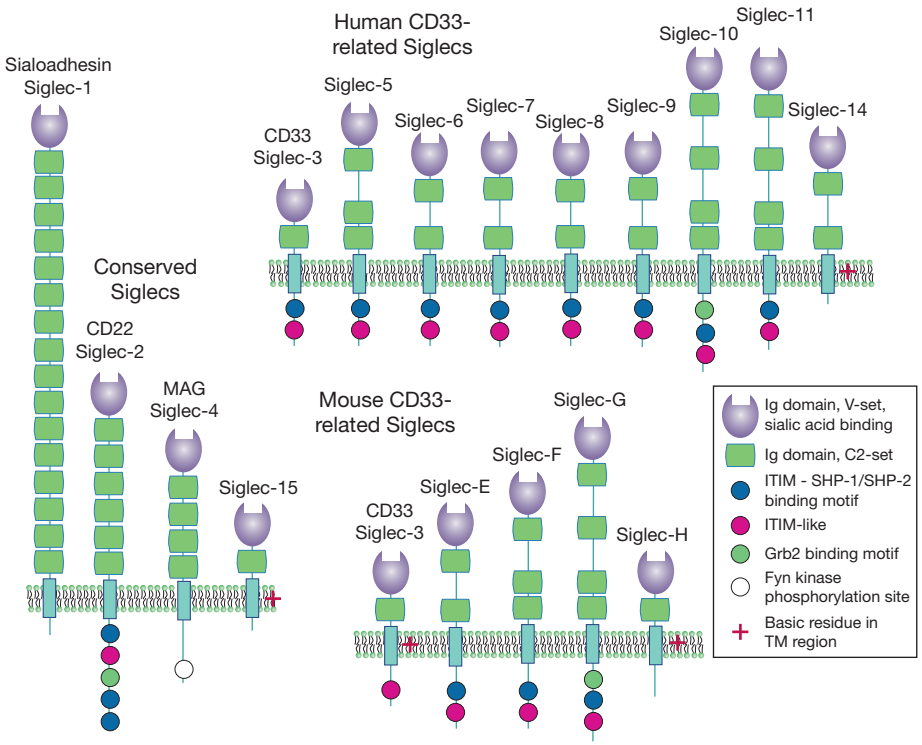

Figure 1 Domain structures of the known Siglecs in humans and mice. There are two main subgroups of Siglecs: one evolutionarily conserved group contains sialoadhesin (Siglec-1), CD22 (Siglec-2), MAG (Siglec-4), and Siglec-15, and the other group contains the rapidly evolving CD33-related Siglecs. In humans, Siglec-12 has lost its arginine residue required for sialic acid binding and Siglec-13 is deleted. ITIM, immunoreceptor tyrosine-based inhibitory motif. The *plus sign* indicates the presence of a charged residue in the transmembrane domain, which has been shown to interact with the immunoreceptor tyrosine-based activatory motif (ITAM)-containing adaptor proteins DAP12 and DAP10. Reproduced with permission from the Consortium of Glycobiology Editors, originally figure 32.1 in Varki A, Cummings RD, Esko JD, et al. (eds.) (2009) *Essentials of Glycobiology*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

most species, the remaining CD33/Siglec-3 related Siglecs appear to have been evolving rapidly. Most of the latter genes are clustered together in an ~500 kb region on human chromosome 19q13.3–13.4 and the syntenic regions of other animals.

Siglec Recognition of Sialic Acids and Their Linkages

The first two Siglecs discovered (Sn/Siglec-1 and CD22/Siglec-2) had strikingly different binding properties for sialosides – with Sn preferring alpha2–3 linked targets and CD22 being highly specific for alpha2–6 linkages. In the latter case, the binding affinity was in the low micromolar range. MAG/Siglec-4 also has an extended binding site that is even highly specific for the underlying sugar chain. There is also variable preference for certain types of sialic acids, with Sn and MAG not tolerating the common *N*-glycolyl modification of Sias. However, the CD33/Siglec-3-related Siglecs are generally more promiscuous in their preferences for different types and linkages of Sias. Of course, many of the less common linkages and types of sialic acids have not been studied for Siglec recognition. The Golgi enzymes that are potential regulators of Siglec functions are primarily the sialyltransferases and the enzymes which modify sialic acids. Some Siglecs also show preferences for certain macromolecular ligands, for example, CD45 and IgM for CD22/Siglec-2, the mucins CD43, and Muc-1 for Sn/Siglec-1, CD24 for Siglec-10/H, and specific brain gangliosides for MAG/Siglec-4.

Potential Effects of Neu5Gc Loss on Human Siglec Biology

The most common Sias of mammalian cells are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc). Humans are a known exception, because of a mutation in cytidine mono-phosphate (CMP)-sialic acid hydroxylase, which occurred after the time (~5–7 Mya) when we shared a common ancestor with great apes. The resulting loss of Neu5Gc and increase in Neu5Ac in humans could have potentially altered the biology of the Siglecs. For example, human cells have a higher density of Sn/Siglec-1 ligands than great apes, the distribution of Sn-positive macrophages in humans is different and a much larger fraction of human macrophages is positive. Other evidence indicates that there are multiple human-specific changes in Siglec biology that may be at least partly related to the loss of Neu5Gc. These include binding changes (in Siglecs -5, -7, -9, -11, -12, and -14); expression pattern changes (in Siglecs -1, -5, -6, and -11); gene conversion (*SIGLEC11*); and deletion or pseudogenization (*SIGLEC13*, *SIGLEC14*, and *SIGLEC16*). Overall, Siglec changes in humans seem to be much more prominent than

in the closely related great apes, suggesting that sialic acid biology was a hot spot in human evolution.

Masking and Unmasking of Siglec-Binding Sites on Cell Surfaces

The initial assumption was that Siglecs were involved in intercellular adhesion. However, in most instances, their binding sites appear to be masked by Sias on the same cell surfaces on which they are expressed. Of course, external ligands with high affinity/avidity can still compete for the endogenous masking ligands. There is also some evidence that unmasking can occur under certain conditions, but it is not known if this is biologically relevant. Overall, the significance of Siglec masking is relatively unclear at this time.

Signaling Motifs in Cytosolic Tails

The inhibitory CD33-related Siglecs (Siglecs -3 and -5–12) have conserved tyrosine residues in the cytosolic tails, one of which corresponds to a canonical immunoreceptor tyrosine-based inhibition motif (ITIM). Various *in vitro* manipulations of these receptors indicate that these tyrosines are indeed targets for phosphorylation and that they can modulate signaling events by recruiting certain tyrosine phosphatases such as SHP-1 and SHP-2. However, the true *in vivo* biological functions of these signaling motifs remain somewhat obscure. Another major unresolved question is: what is the connection between extracellular sialic acid recognition and signaling via the cytosolic motifs? Even less is known about the functions of the more recently described activatory Siglecs (Siglecs 13–16), which recruit the DNAX-activating protein (DAP) 12 adaptor along with its canonical immunoreceptor tyrosine-based activating motif (ITAM). It appears likely that these are balancing receptors that generate opposing inhibitory and activatory responses, perhaps in response to sialic acid-expressing pathogens.

Known and Putative Functions of the Siglecs

Various lines of evidence indicate that MAG/Siglec-4 is involved in the maintenance of myelin organization and in the inhibition of neurite outgrowth during regeneration after injury. It is also clear that CD22/Siglec-2 functions as an inhibitory component of the antigen receptor complex of B Cells, and is thus involved in regulating the humoral immune response. While Sn/Siglec-1 appears to mediate various macrophage adhesion events *in vitro* and *in vivo*, it is still somewhat unclear what exactly the functions of these interactions are. This molecule may also assist macrophages in the phagocytosis of sialic acid-expressing pathogen. Relatively little is known about the functions of CD33-related Siglecs. It appears that these molecules are involved in innate immunity. One hypothesis supported by current data is that Siglecs are sensors for pathogens that have sialylated cell surfaces and/or express extra cellular sialidases. The corresponding endogenous function of the inhibitory CD33-Siglecs appears to be to recognize host Sias as self and dampen unwanted innate immune reactivity. Several findings support this notion. For example, mice deficient in Siglec-F (which is normally found on eosinophils) show hyper-eosinophilia and exaggerated allergic responses, and group B Streptococci engage neutrophil Siglec-9 via their sialylated capsules to dampen innate immune responses. Also, human T and B cells express lower levels of CD33-related Siglecs than in great apes and this correlates with increased responses of the human cells to stimuli. It is suggested (but not proven) that the activatory CD33-related Siglecs (13–16) have opposing effects, via their cytosolic ITAM motifs.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Lectins; **Signaling:** Immunoglobulin (Fc) Receptors.

Further Reading

Crocker PR, Clark EA, Filbin M, et al. (1998) Siglecs: A family of sialic-acid binding lectins [letter]. *Glycobiology* 8: v2.

Crocker PR, Mucklow S, Bouckson V, et al. (1994) Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. *EMBO Journal* 13: 4490–4503.

Crocker PR, Paulson JC, and Varki A (2007) Siglecs and their roles in the immune system. *Nature Reviews Immunology* 7: 255–266.

Kelm S, Pelz A, Schauer R, et al. (1994) Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Current Biology* 4: 965–972.

Powell LD, Sgroi D, Sjoberg ER, Stamenkovic I, and Varki A (1993) Natural ligands of the B cell adhesion molecule CD22beta carry N-linked oligosaccharides with alpha-2, 6-linked sialic acids that are required for recognition. *Journal of Biological Chemistry* 268: 7019–7027.

Powell LD and Varki A (1995) I-type lectins. *Journal of Biological Chemistry* 270: 14243–14246.

Varki A (2010) Colloquium paper: Uniquely human evolution of sialic acid genetics and biology. *Proceedings of the National Academy of Sciences of the United States of America* 107(supplement 2): 8939–8946.

Varki A and Angata T (2006) Siglecs – the major subfamily of I-type lectins. *Glycobiology* 16: 1R–27R.