Harnessing the power of AI for sialobiology research and education

Arun K. Datta

National University, San Diego, CA (USA)

Since the sialic acid was first discovered by Gunnar Blix in 1952, a lot of information has been gathered regarding its role in the development and diseases. This class of alpha-keto acid sugars with a nine-carbon backbone, commonly found in animal tissues, has always drawn the interests of sialobiologists because of its multiple interesting cellular functions. This molecule is synthesized via a multi-step process involving several enzymes and precursor molecules. The activated form, predominantly N-acetylneuraminic-9-P, is then utilized for various cellular functions including cell signaling and cell-cell Interactions. The most common forms of sialic acid found in primates are N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc). However, due to mutation in the CMAH gene, humans cannot synthesize Neu5Gc, thus distinguishing us from other primates. This difference has implications for various biological processes, including our immune response and the way we are affected by certain pathogens. There are over 50 different isoforms known for this molecule.

Sialic acid primarily resides on the cell surface, forming conjugates with glycoproteins, glycolipids, and gangliosides. Sialylation of these glycoconjugates occurs predominantly by enzyme mediated reaction utilizing its activated form, CMP-Sialic acid. A family of sialyltransferases are involved in this enzymatic transfer, which occurs in the Golgi apparatus of the cell, where glycosylation reactions are completed before the glycoconjugates are transported to the cell surface or secreted out of the cell. It can also be transferred by trans-sialidase as is found in some parasites, such as in *Trypanosoma cruzi*. Sialic acid is also found in certain other organisms. Moreover, some bacteria and viruses can uniquely use this molecule from the host to camouflage themselves against the immune system. The specific type of linkage formed and the position of sialic acid in the glycan chain can significantly influence the biological function of these glycoconjugates.

More information about this fascinating molecule can now be gathered quickly by using generative AI, such as ChatGPT from OpenAI, Copilot from Microsoft, or Gemini from Google. Since the groundbreaking paper “Attention Is All You Need” by Vaswani, et. al. (2017)1 of Google BRAIN, a lot of interest has been generated in this field of machine learning. Their seminal work introduced the Transformer model, which is based solely on attention mechanisms. Application of this technology opens up the possibility of developing the GlycocodeNetwork, similar to MotifNetwork, for better understanding its role in cellular interactions. This presentation will demonstrate this technology and explain the fundamental aspect of generative AI and its various models. This presentation will also include the demonstration of Glycomics Workbench2 that utilizes this powerful technology for Glycome research and education.

1. Attention Is All You Need. Vaswani, A., Noam Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, N. Kaiser, L., Polosukhin, I. (2017). NIPS 17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 2017, Pages 6000–6010.

2. Glycomics Workbench, a grid technology-based workbench for Glycome analysis. Arun K Datta and Nitin Sukhija (2020). Proceedings in the 13th annual NIH & FDA Glycoscience Research Day, Bethesda (Maryland), May 15, 2020.