Tuesday

Software

#4

Glycomics Workbench: Harnessing the Power of AI for Deciphering the Glycocodes

Arun Datta[1], Nitin Sukhija[2] and Kimberly Bruch[3]

[1]National University San Diego, CA, USA
[2]Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA
[3]University of San Diego, SD Supercomputer Center, San Diego, CA, USA

Glycans, more specifically, glycocodes [1], which are complex carbohydrate structures consisting of ten monosaccharides in humans, encode information for various biological processes, including cell-cell interactions, extracellular signals, and cell differentiation. Aberration in the structures of these glycocodes have been observed in various pathological conditions including cancer. As our understanding of glycoscience grows, researchers increasingly recognize the significance of glycocodes in development and diseases, therefore, attracting attention from biotech/pharma companies aiming for biomarker discovery and developing carbohydrate-based drugs. However, deciphering glycocodes presents unique challenges. Glycomics, the structure-function study of glycocodes, are difficult to perform compared to other 'omics' fields. Unlike the linear flow of information in the Central Dogma, Glycomics require detailed and meticulous analysis of glycosyltransferases, and glycosidases that orchestrate the assembly and modification of glycans, resulting in diverse glycocodes.

Glycan-based drug discovery faces limitations due to the complexity and intricate nature of glycans—varying in size, branching patterns, and modifications, thus, posing hurdles for analysis and current detection technologies. Moreover, manual curation for generating information for learning and conducting research on a glycocode's structure and function is time consuming. Recent developments in the machine learning technologies, particularly GenAI, has created an unprecedented opportunity that can tremendously help in such analysis and better understanding the complexity involved. GenAI can also accelerate glycocode analysis by automating tasks that once required weeks or even months of manual curation. Similar machine learning technologies, such as, AlphaFold and Rosetta, can be exploited to create artificial enzymes for chemoenzymatic synthesis of glycocodes for therapeutic purposes. Glycomics Workbench is designed to achieve these goals.

This CI- supported portal [2] is under development with 'Molecule page' defining every glycan related molecule. This presentation will include showing its features including the demonstration of the web-based Molecule page for ST6Gal I [3] as an example.

References

[1] Datta, AK., and Sukhija, N. (2021). Glycobioinformatics in deciphering the Mammalian Glycocode: Recent Advances. In "Glycome: The Hidden Code in Biology" (D. Banerjee, ed.) Nova Science Publishers, article 16 (pages 323 - 376).

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

[3] Tsuji, S., Datta, AK., and Paulson, JC. (1996). Systematic nomenclature for sialyltransferases. *Glycobiology*, **6** (7)1:v-vii