From Static to Dynamic: Integrating Glycan Dynamics into Structural Biology

Proteins are the molecular machines that power life, controlling everything from how cells function to how diseases develop. Their 3D structures are critical for understanding how they work and for developing new drugs and therapies. Structural biology techniques have revealed over **220,000 protein structures to date**. These data fuel biological sciences, including **AI tools** like **AlphaFold**, which can predict protein structures from their amino acid sequence. However, methods used to solve structures rely heavily on averaging information from many copies of proteins, which makes flexible parts of proteins blurred or completely invisible. This is particularly true for **glycans**—sugar molecules attached to most proteins. Glycans are crucial for cell communication or immune responses, but because they are highly mobile, they are often incomplete or completely missing in structural databases; similar situation also occurs for other flexible protein parts. This leaves us with an **incomplete understanding of protein function**, limiting **drug design** and decreasing the quality of **AI predictions**.

Our project aims to redefine how we look at protein structures by incorporating **dynamic ensembles of conformations**, focusing on flexible regions like glycans. Even though these regions are difficult to capture at high resolution, existing low-resolution data contain valuable information about the range of shapes they adopt. Combining **molecular dynamics simulations**, our glycan modeling tool **GlycoSHIELD**, and **machine learning with existing structural biology datasets**, we will model the full, flexible structures of proteins, starting with medically important proteins like **cellular receptors** and **viral glycoproteins**.

We will then apply these insights to **whole proteomes**, incorporating **tissue- and disease-specific glycosylation** data. This will allow us to model how protein structures change in different tissues and diseases, such as **cancer** and **neurodegenerative disorders**. By capturing the **true, dynamic nature of proteins**, this project has the potential to **transform drug design**, **vaccine development**, and **AI predictions**, offering new opportunities for **targeted therapies**.