Abstract

Glycosaminoglycans (GAGs) are linear anionic polysaccharides composed of repeating disaccharide units made up of a hexosamine and a uronic acid with varying degrees and patterns of sulfation. They are essential components of the extracellular matrix and are pivotal in numerous cellular processes by interacting with various proteins. Disruptions in these interactions can cause severe health conditions including neurodegenerative and inflammatory diseases.
The specificity of interactions between GAGs and proteins, i.e. the ability of the receptor (protein) or ligand molecule (GAG) to distinguish between similar targets and preferentially form interactions with one or a few selected binding partners, is largely governed by the structural and chemical properties of GAGs, including their sulfation patterns, charge distribution, length, and flexibility, as well as the protein's structural features. Disruptions in binding specificity caused by protein mutations, alterations in GAG structure, or changes in the cellular environment can result in protein dysfunction, dysregulated signaling, and disease. Therefore, understanding binding specificity is crucial for developing GAG mimetics for therapeutic use to recreate, enhance, or modify natural GAG binding.
Isolating structurally homogeneous fragments of unbound GAGs is difficult due to the variations in their chain length, sugar composition, and sulfation pattern. Once GAGs bind to proteins, their increased conformational flexibility and transient interactions become the chief obstacles to structural analysis. The dynamic multivalent nature of protein/GAG interactions, often driven by electrostatics, can result in multiple binding poses with comparable affinities, which complicates the detailed characterization of specificity of individual binding sites *in vitro*. A combination of experimental and computational methods is beneficial to address these challenges and provide a thorough understanding of protein/GAG specificity.
Computational methods elucidate binding mechanisms and the effects of specific molecular features of GAGs at the molecular level. Mapping the electrostatic potential using theoretical models helps to identify protein regions that are likely to interact with GAGs. Molecular docking and molecular dynamics simulations

are used to identify potential binding sites and poses, and to characterize interaction stability and conformational changes within the protein/GAG complex. All-atom simulations provide atomistic detail while coarse-grained simulations are able to capture larger scales over extended timescales owing to the grouping of atoms into "beads". By quantifying the energetic cost or gain of binding, free energy calculations provide quantitative insight into the thermodynamic favorability of protein/GAG binding. Other approaches, such as machine learning, can classify interactions based on the structural properties of both proteins and GAGs, revealing hidden patterns in their binding mechanisms and highlighting key features driving binding specificity.

The research presented in this thesis aimed to explore areas currently beyond experimental reach in the study of protein/GAG binding specificity using computational methods. The employed computational models were integrated with and validated against experimental data to provide a comprehensive, multi-scale understanding of protein/GAG interactions. Using Poisson–Boltzmann calculations, I quantified how different GAGs affect the protein's near-surface electrostatic field and validated these changes by comparing the computed maps with measurements from a novel paramagnetic-probe NMR technique. Next, I performed atomistic molecular dynamics simulations of protein/GAG and protein/peptide complexes to explore the influence of electrostatic interactions versus other physicochemical properties on binding specificity. Free energy calculations decomposing the binding energy into separate contributions and analysis of ligand flexibility distinguished GAG interactions from acidic peptides. Furthermore, I implemented, fine-tuned, and validated a coarse-grained model of heparin. Finally, I performed MD simulations of GAGs with different sulfation patterns in combination with unsupervised learning to establish links between GAG structural features and binding specificity.

By utilizing computational chemistry techniques, the research presented in this thesis made key contributions to understanding the specificity of protein/GAG interactions. Innovative tools and analysis pipelines were designed for coarse-grained simulations, enhancing the efficiency of protein/GAG interaction modeling. This research demonstrated the complementary nature of computational and experimental methods in studying protein/GAG complexes. The developed computational approaches support large-scale simulations and offer a robust framework for future studies on protein/GAG binding specificity relevant to therapeutic design.