**Computational Approaches for Mass**

**Spectrometry-based Characterization**

**of Antibody Repertoires**

Scientific summary

Antibodies are crucially important molecules with a key role in the human immune system. Their pivotal role in combatting infections and other diseases has led to recombinant therapeutic antibodies taking center stage for development of therapies in the last decade. Consequently, there is a growing demand for efficient and high-throughput characterization methods to streamline discovery of new antibodies and development pipelines.

A fundamental aspect of protein characterization is determining its amino acid sequence. Traditionally, antibody sequencing has relied on B-cell receptor sequencing, which provides information about the genetic blueprint of the antibodies. However, there is a growing interest in protein-level sequencing of antibodies to directly analyze circulating antibody repertoires and the relative clonal abundance in specific biological contexts. Mass spectrometry (MS) has emerged as a promising method for protein sequencing and quantitative analysis. However, antibodies represent a uniquely challenging class of proteins. Estimates suggest that the human body can generate over 10^15 unique antibodies. These antibodies all have unique yet highly similar sequences, posing a tremendous challenge for their characterization and sequencing. Thus, while MS has already been used to characterize and sequence highly purified monoclonal antibodies, the field has not yet moved towards antibody sequencing from endogenous, polyclonal samples, highlighting the need for further refinements in sample preparation and data analysis to fully harness the potential of MS in antibody sequencing and repertoire analysis.

This thesis describes the development of computational strategies for the analysis of endogenous antibody repertoires. At the repertoire level, we show how longitudinal quantification of individual antibody clones can elucidate repertoire dynamics in response to physiological events such as vaccination or disease. Using advanced intact protein MS techniques, we were able to detect and quantify immunoglobulin molecules in human serum and breast milk and construct personalized IgG1 and IgA1 repertoires. These repertoires were highly personalized but surprisingly simple, with a limited number of clones (several hundreds) dominating each repertoire. Additionally, while repertoires of healthy individuals were longitudinally stable, significant changes in repertoire composition were observed following immune challenges.

We also sequenced endogenous antibody chains directly from serum by leveraging the synergy between peptide- and protein-centric MS approaches. The peptide-centric approach provides comprehensive coverage of the antibody sequences, while the protein-centric fragmentation approach offers chain-specific sequence information. In addition to an initial proof of concept, where we sequence an abundant clone from serum, we also present a more generalizable workflow, representing a crucial step towards automation of the process, validated by sequencing a series of increasingly complex samples.

We conclude that MS can be used to monitor circulating antibody dynamics and sequence endogenous antibody clones from serum. Through antibody profiling, we can gain new insights into the generation, timing, and specificity of antibody responses to physiological stimuli, which could be used to select potential therapeutic candidates directly from serum. Directly sequencing these candidates using MS presents an exciting opportunity to facilitate drug discovery by simplifying and accelerating therapeutic development pipelines.