**Chapter 1**

**1**

## Thesis overview

Throughout this thesis, I detail my efforts to develop computational workflows and tools that facilitate the analysis of complex LC-MS data. There is a strong focus on antibody repertoires, apart from Chapter 2 which focuses on cross-linking MS data. The work described in this chapter shaped what became the guiding principle of my academic efforts; that *standardized computational tools are of vital importance for reproducible research.* As such, I consider it the spiritual predecessor to the subse­ quent chapters and an important example of the central theme.

**Chapter 2** describes how we developed CrosslD, a tool to analyze large and com­ plex cross-linking proteomic datasets. CrosslD was developed to facilitate explo­ rative analysis of large amounts of crosslinking data. We show that integration of data from multiple sources can provide valuable insights, as the integrated data from protein databases enables gene ontology enrichment analysis and grouping based on function. Furthermore, we showcase how mapping of crosslinked residues onto 3D-structural models for proteins can help refine these models or help to generate models for protein complexes.

In **Chapter 3,** the antibody repertoire profiling technique which enabled the research in **Chapter 4** and **5** is introduced. In this initial application of the technique on a co­ hort of sepsis patients we found the serological lgG1 repertoires to be unique to each individual, stable over time, responsive to physiological events and relatively simple, consisting of several hundred clones despite there being an enormous number of theoretically possible clones. Furthermore, this chapter provides proof of concept for *de nova* sequencing of endogenous antibodies by using a multi-tier mass spectrom­ etry approach to sequence the most abundant clone for a donor.

**Chapter 4** describes the analysis of breastmilk SlgA1 profiles of six mothers who had received two identical SARS-CoV-2 vaccinations over 16 timepoints. We use

#### Chapter 1

**1**

the extensive sampling and repeated vaccination to define clonal populations based on the detection window of these clones relative to the vaccination moments. We also discover that the second vaccination induces the emergence of a population of novel clones and show that titer fluctuations as measured by ELISA can be driven by highly divergent clonal populations.

In **Chapter 5,** we build upon the proof of concept for *de nova* sequencing of en­ dogenous antibodies by hybrid mass spectrometry. We present a more standard­ ized workflow for sequencing antibody chains in mixtures. Our approach resolves ambiguity in sequence predictions for the hypervariable complementarity determin­ ing regions by mass-filtering candidate sequences based on the gap size between adjacent framework regions, which we determine using middle-down fragmentation data.

**Chapter 6** contains a summary and a discussion of the advances that enabled the work in this thesis, the impact of the findings for others in the field, the challenges that lay ahead and how they may be overcome, along with an outlook on where I believe the field is heading.

Finally, **Chapter 7** is a tribute to all the amazing people without whom this research and thesis would not have been possible.