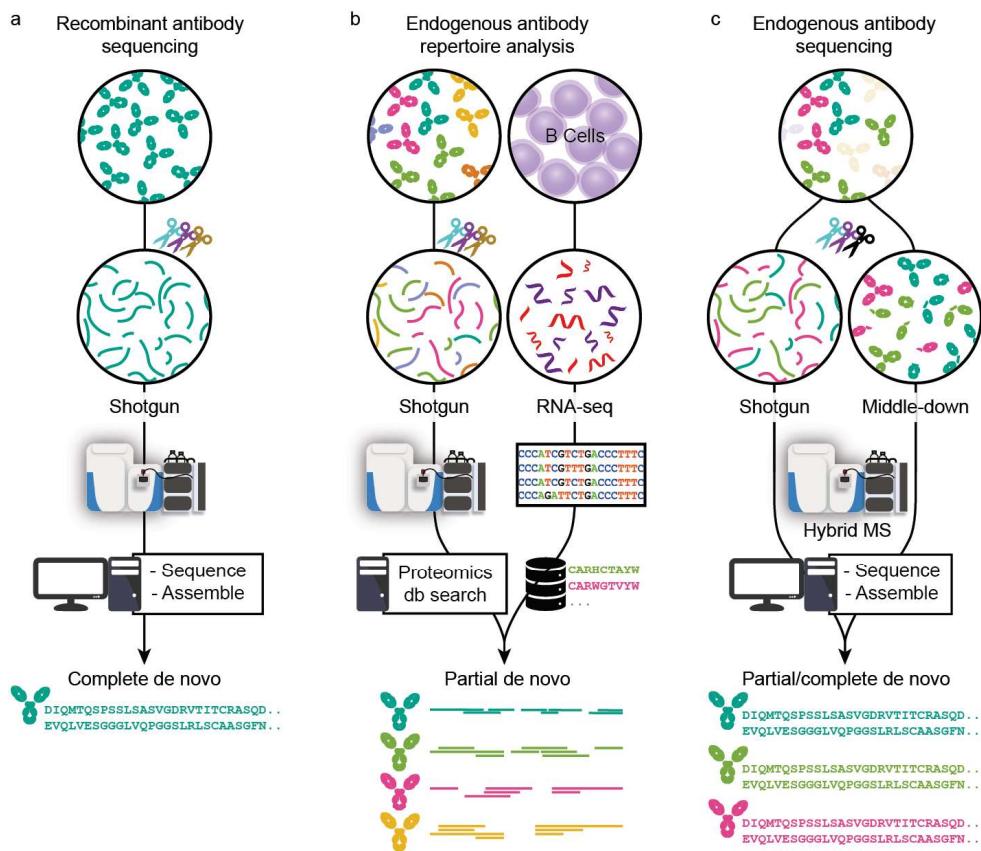


Introduction



This chapter includes parts of the following publication:

A perspective towards mass-spectrometry-based de novo sequencing of endogenous antibodies

Sebastiaan C. de Graaf*, Max Hoek*, Sem Tamara and Albert J.R. Heck

mAbs (2021), 14:1, 2079449, DOI: 10.1080/19420862.2022.2079449 Review

* These authors contributed equally to this work

1.1 Prelude - The importance of antibodies

AROUND the time of their initial discovery, antibodies were termed by various illustrious names, such as ‘Immunkörper’, ‘Amboceptor’, and ‘Zwischenkörper’, among many others. These terms were used more than a century ago to describe substances with antitoxin, lysis, agglutinin, and precipitin activities.(1, 2) Nowadays, the generally accepted term *antibody* refers to secreted immunoglobulins (Igs), whose sequence variety is several orders more diverse than the assortment of their historical names. Antibodies represent some of the most important molecules in the human immune system. Over the last century, Igs have been intensively studied because of their role in combatting infectious diseases and have taken centre stage for development of therapeutics in the last decade.(3–5) Beyond infectious diseases, recombinant antibodies are now also developed for cancer, rheumatoid arthritis, and various other pathological conditions.(6) As key entities in the body’s defence mechanism, circulating antibodies are found in various bodily fluids, such as serum, saliva, milk, the lumen of the gut, and cerebrospinal fluid.(7) New leads for biotherapeutic development of recombinant antibodies come either from immunizing animals with specific antigens, or by discovering pathogen-neutralizing antibodies from recovered patients.(8–10) The estimated diversity of Ig molecules a human body can generate extends beyond 10^{15} theoretical sequences,(11, 12) indicating that each antigen may lead to a unique antibody response. These 10^{15} possible antibody sequences are all unique yet highly alike, posing a serious challenge for their characterization and sequencing, which has remained, to this day, a tremendously challenging task. Ideally, one would like to sequence antibodies at the protein level instead of through B-cell receptor (BCR) sequencing,(13) as is currently the norm, to more directly probe circulating antibody repertoires and their relative abundances in specific environments. Mass spectrometry (MS) is expected to be the method of choice to potentially achieve this feat, as MS-based protein analysis has advanced and matured considerably.(14, 15)

However, antibodies represent a very special and rather challenging class of proteins. Consequently, while MS has been already used to characterize and sequence highly purified monoclonal antibodies (mAbs),^(16–18) further technical developments in sample preparation and data analysis are needed to incorporate MS fully and efficiently into an endogenous humoral antibody discovery and characterization pipeline. In this thesis, I evaluate the role that MS can play in sequencing, identifying, and characterizing antibodies, focusing mainly on emerging strategies employed to enable identification and characterization of endogenous neutralizing antibodies.

1.1.1 Nomenclature, structure, and diversity of antibodies

Humoral human antibodies are complex proteins produced by B cells.^(7, 19) Most antibody molecules (e.g. IgGs) are made up of four protein chains: two identical light chains and two identical heavy chains, which are interconnected by disulphide bridges (**Figure 1**). The light and heavy chain form two heterodimers, which are connected via disulphide bridges in the hinge region to form the intact antibody. Functionally, the intact antibody can be divided into two antigen-binding domains (also known as Fab or fragment antigen-binding) and a constant domain (also known as Fc or fragment crystallizable)⁽²⁰⁾ (**Figure 1a**). The Fc is the effector entity of the antibody and can bind to Fc-receptors on immune cells⁽⁷⁾ and mediate immune effector responses such as phagocytosis, antibody-dependent cell-mediated cytotoxicity, respiratory burst, and cytokine release.⁽²¹⁾ In contrast to the fully conserved sequence and structure of the Fc, the Fab is responsible for the vast diversity in recognized antigens and is thus hypervariable. Because there is an endless and constantly evolving pool of pathogens, the antibody repertoire needs to be incredibly diverse and versatile to counteract these challenges.^(22, 23) In humans, this enormous diversity in the potential antibody repertoire is achieved through several mechanisms. Starting at the genomic level, the light and heavy chains are encoded in four genes each:

Variable (V), Diversity (D), Joining (J), and Constant (C), with the light chain lacking the D-gene. These genes are encoded in multiple alleles, which can recombine to a staggering number of combinations (**Figure 1b**).⁽²⁴⁾ The recombination process is also error-prone, leading to insertions and deletions at the junctions between the regions, referred to as junctional diversity. By recombination alone, the number of possible variable domain sequences already reaches tens of thousands. However, the eventual antibody diversity is expanded even further by natural polymorphisms, mutations, and class switching. As the major contributor to antibody hypervariability, somatic hypermutations can occur during B-cell affinity maturation and do so at a million-fold increased rate compared to the usual mutation rates.⁽¹¹⁾ These mutations are largely concentrated in the complementarity-determining regions (CDR1-3), separated by framework regions (FR1-4), which form the conserved backbone of the Fab structure (**Figure 1c**). Located at the tips of the Y-shaped antibody structure, CDRs are primarily responsible for antigen binding, and, therefore, elucidation of their sequences is of the utmost importance for antibody discovery. The Fc part of Igs is used to classify antibodies into one of 5 classes: IgA, IgD, IgE, IgG, and IgM. Some of these classes are divided further into subclasses denoted by numbers, e.g., IgG1-4 or IgA1 and IgA2. Although the function of the classes and subclasses is different, their variable regions stem from the shared pool of genes. Therefore, for simplicity, in this review, we focus primarily on IgG1, the most abundant antibody subclass in serum, and the predominantly used subclass for biotherapeutic development. Still, concerning *de novo* sequencing by MS, different Ig classes and subclasses pose similar challenges and opportunities.

1.1.2 Modalities of MS-based antibody analysis

Proteomics is the large-scale study of proteins. Many different peptide- and protein-centric MS-based approaches have been developed for proteomics, whereby some

Chapter 1

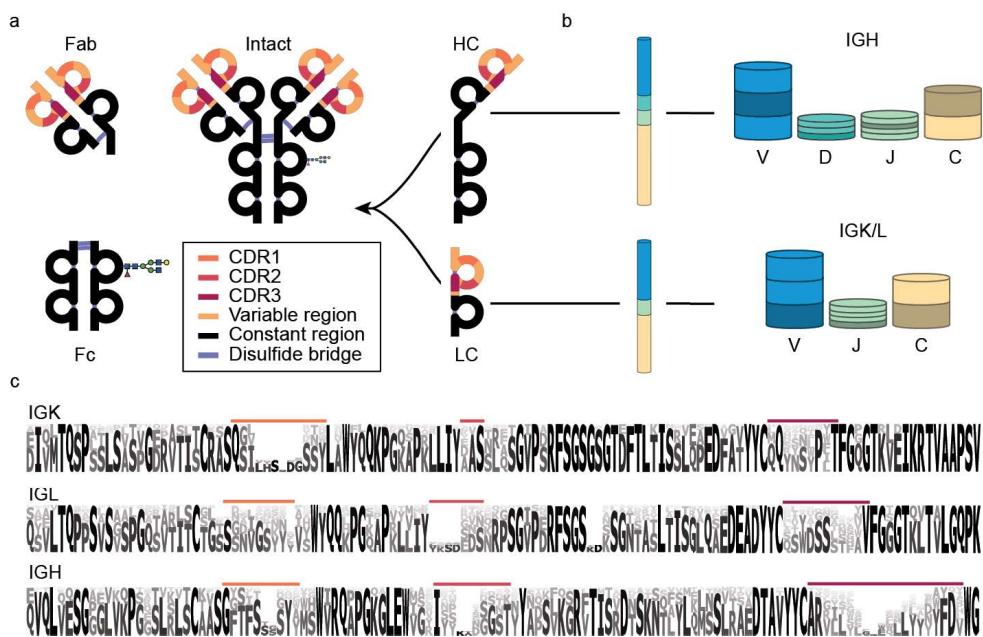


Figure 1: Nomenclature, structure, and diversity of IgG1 antibodies. (a) Nomenclature and protein fragments of an IgG1 molecule. The antigen-binding domain, containing light and heavy chain (LC and HC respectively) variable regions, is termed Fab (or Fab2 when dimerized). The constant part of the heavy chain carrying an N-glycosylation site is called Fc. Other IgG subclasses vary in their heavy chain constant region (Fc) and disulphide patterns. (b) The diversity in antibodies originates primarily from the V, D, J, and C-allele (each annotated with a distinct colour) recombination process. In this process, each of many individual V, D, J, and C-alleles can recombine with any of the other gene segments, yielding thousands of possible combinations, in particular for the heavy chain, which incorporates the most diverse D region. (c) Sequence logo created by the alignment of in silico generated sequences of Ig kappa (IGK) and lambda (IGL) light chains and Ig heavy chain (IGH) from the international ImMunoGeneTics (IMGT) information system database.(25) Even though the displayed sequences are part of the variable domain, large stretches of these sequences, also known as the framework regions (FRs), are relatively conserved, compared to the hypervariable complementarity determining regions (CDRs), coloured in accordance with (a).

of these have been adapted for de novo sequence analysis of antibodies. Bottom-up (BU) or shotgun proteomics is by far the most widespread approach in MS-based protein analysis. In it, protein samples are digested by one or more proteases, and the resulting peptides are separated by some form of liquid chromatography (usually reversed-phase (RP)-HPLC), after which their peptide masses are recorded (MS1). Highly abundant precursor ions are then selected for fragmentation, and the masses of their fragment ions (MS2) are recorded. Because digestion and MS-based fragmentation adhere to highly specific rules, peptides and their gas-phase fragment ions can be predicted. Consequently, peptides and their parent proteins are iden-

tified by comparing recorded spectra to the spectra simulated from protein or DNA databases.(26) For antibody sequencing, personalized databases are required for identification. Yet, digestion-based strategies are still widely used even without an available database. Individual spectra can be *de novo* sequenced, and the resulting reads can be assembled into full-length sequences.(16, 27, 28) Additionally, intact mass analysis is a useful tool for protein analysis, providing masses that can be considered fingerprints of the species (known as proteoforms) present in the sample. Comparing different masses can lead to conclusions about relations between multiple species, for instance, if they differ by the mass of a known mutation, post-translational modification (PTM), or signal peptide.(29) In the case of antibodies, such analysis can be performed with the protein in its native, and possibly complexed, state, or unfolded and separated into the comprising chains. Such approaches can provide valuable insights in the context of antibodies, e.g., by assessing the complexity of antibody repertoire or following changes in abundance of specific clones.(30) When applied to *de novo* sequencing, the precursor mass knowledge can help determining the light and heavy chain pairing or sequence prediction accuracy in BU sequencing.(31) In addition, both denatured and native antibodies can also be fragmented to yield some sequence information, this approach is called top-down (TD) MS. Because of the much larger size and higher charge of the analysed species, such intact-protein fragmentation spectra are more complex and harder to interpret than peptide spectra.(32, 33) To mitigate this, specific proteases can be used to cleave proteins into smaller subunits. This practice is called middle-down (MD) MS, and in the context of antibodies it is often performed by cleaving the hinge region of the heavy chain before MS analysis.(34) Fragmentation spectra of entire chains or intact antibodies can provide valuable tools for both sequence determination and validation of sequence predictions, as fragmentation is highly specific for the precursor clone, which is often untrue in BU analysis.(35)

1.2 The emerging role of mass spectrometry in antibody discovery

Due to the structural complexity and immense sequence diversity of antibodies, the development of therapeutic antibodies has always been a very challenging and labour-intense task, especially when compared to small-molecule drug development. For example, the discovery of Trastuzumab was achieved by using mice immunized with antigen-expressing cells. Following the generation and selection of hybridomas that showed specific activity,(36) the sequence of the selected antibody was determined after cloning and expression. A humanized antibody could be produced only thereafter by adapting and modifying the sequence accordingly.(37) The same approach was used in the development of several other mAbs.(38–41) Apart from being expensive and laborious, these early strategies required knowledge and availability of purified antigens and animal models that can produce specific antibodies in response to these antigens.(42) More recently, alternative strategies for antibody discovery have been explored starting with the screening of B cells from individuals who successfully overcame an infection. In this approach, peripheral blood mononuclear cells (PBMCs) are isolated, immortalized, and screened for antigen reactivity. The reactive clones are further expanded and characterized. This method has proven effective in finding new neutralizing antibodies that can be used to combat certain infectious diseases, e.g. Ebola(8, 9) or severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).(10) These recent advances show that the discovery of antibodies from human subjects, in addition to animal models, represents a viable method for developing new avenues for therapies. However, it may be even more advantageous to discover and characterize mature antibody clones directly from clinical samples at the protein level in their functionally matured and active form.

In recent years, MS-based proteomics has advanced tremendously in sample preparation, MS and liquid chromatography instrumentation, and data analysis.(14, 15)

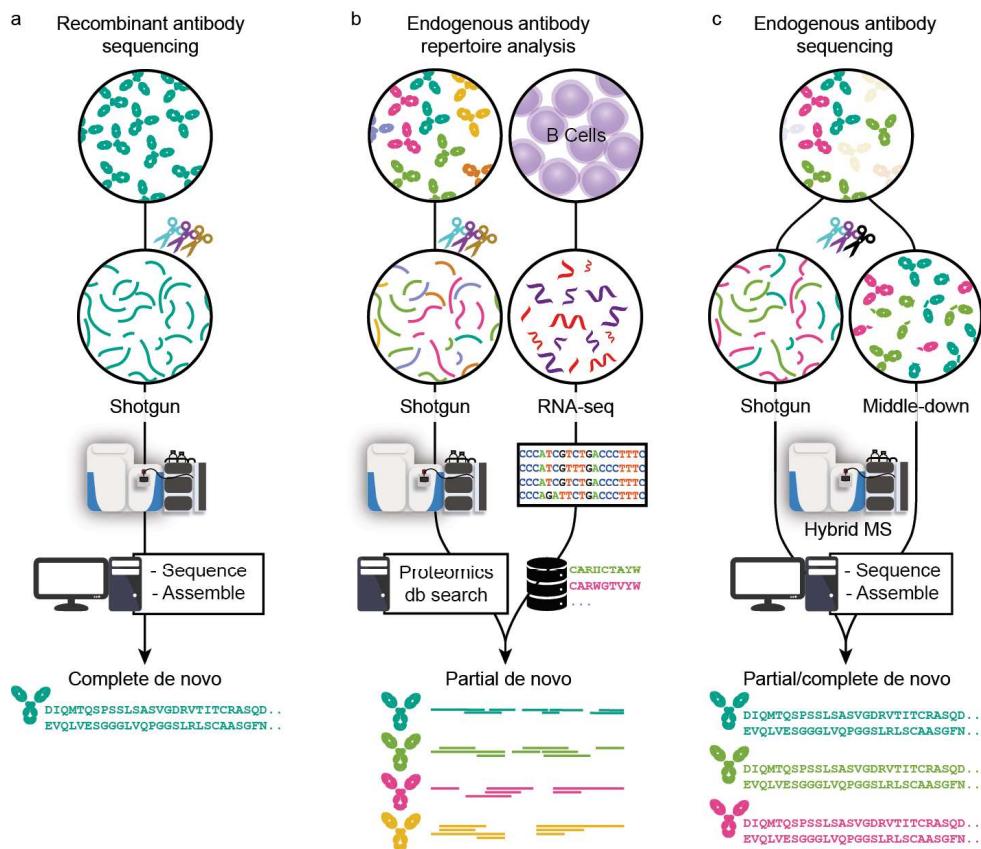


Figure 2: Three approaches in MS-based antibody sequencing. (a) Recombinant antibody sequencing generally starts with abundant highly purified mAbs, which can be fully sequenced through BU MS, where hundreds of peptides are generated by digesting the mAb with one or several proteases, providing multiple overlapping short sequence reads. After liquid chromatography-mass spectrometry (LC-MS) measurement, the spectra can be processed by several different de novo sequencing software solutions and assembled into full-length mAb sequences.(17) (b) In repertoire analysis, a sequence database is generated through B-cell sequencing, and MS-data is obtained through BU MS experiments. After generation of the personalized database, a high throughput of the LC-MS is possible.(43) While not strictly de novo since only hits from the sequence database are identified, it is a powerful tool for antibody repertoire analysis. (c) Endogenous antibody sequencing cannot rely on BU MS alone, as direct sequencing of endogenous humoral antibodies is hampered by inherent challenges and complexity. Emerging MD and TD MS techniques provide clone-specific sequence information highly complementary to traditional sequencing. Integrating BU MS and MD/TD MS makes it possible to achieve full-length coverage of antibody sequences.(30)

Using all these advances, antibody sequencing at the protein level by MS has come within reach. **Figure 2** highlights three – chiefly MS-based – strategies used to determine antibody sequences. These three pillars are primarily classified by the source of sample material and the attainable sequencing information. The first strategy applies to highly purified recombinant antibodies that are now amenable for full se-

Chapter 1

quencing with BU MS, often by combining several different proteases and advanced algorithms. Second, hybrid approaches have been introduced for analysing endogenous antibody repertoires by combining MS-based techniques with genomics or transcriptomics, e.g., whole genome sequencing or BCR sequencing, ideally from the same donor. The third set of techniques encompasses several MS-based *de novo* approaches that aim to determine complete antibody sequences of selected clones directly from clinical samples without the aid from alternative omics data. While each strategy is distinct, they all share common aspects.

1.2.1 MS-based sequencing of monoclonal antibodies

Before delving into the topic of MS-based sequencing of endogenous antibodies from clinical samples, we first discuss the current state-of-the-art sequencing approaches developed for recombinant mAbs. Principles of mAb sequencing by MS share many technical considerations with sequencing of antibodies present in complex mixtures. Furthermore, currently available strategies for recombinant mAb sequencing provide great context for discussing limitations and bottlenecks that hamper sequencing of endogenous antibody clones.

Shotgun, bottom-up strategies used for sequencing of highly purified mAbs

Antibodies are often analysed after digestion with one or more proteases to generate peptides (**Figure 2**). Such peptide-centric approaches are known as BU MS and represent the most popular type of proteomics experiments. In contrast to most shotgun proteomics experiments, *de novo* sequencing through BU MS necessitates a high depth of sequence coverage, i.e. each sequence position in the antibody is ideally supported by multiple overlapping unique peptides. With a typical highly specific protease such as trypsin, which cleaves C-terminally of lysine and arginine, and a low number of missed cleavages, sequence-coverage depth is often limited because only

a few of the generated peptides overlap in sequence. Although this suits standard shotgun proteomics experiments, which do not require full sequence coverage of the analysed proteins, *de novo* sequencing of antibodies thus requires other approaches. Several methods to generate complete and deep sequence coverage by overlapping peptides have been introduced. For example, a shortened protease incubation time was successfully used to increase the number of peptides carrying missed cleavage sites.(44) Some proteases generate a high number of overlapping peptides through non-specific cleavage.(31, 45, 46) Alternatively, non-specific cleavage can be also achieved through non-enzymatic treatment, e.g., microwave-assisted hydrolysis.(47) For these methods to work, digestion conditions must be tightly controlled to avoid abnormally long or short peptides and ensure reproducibility. Another elegant option is to use multiple proteases with synergistic sequence specificities. For instance, Peng et al.(17) recently used a total of 9 proteases, both specific and non-specific, to successfully *de novo* sequence a full-length anti-FLAG-M2 mouse mAb (**Figure 3**). The strength of a large panel of proteases is shown in the validation of the CDR sequences by high scoring peptides covering the entire CDR. The 6 chosen peptides are the result of digestion by 5 different proteases (trypsin, chymotrypsin, lysC, thermolysin, and elastase, **Figure 3a and b**). Nowadays, most *de novo* sequencing solutions, such as ALPS/PeaksAB,(27) GenoMS,(46) SuperNovo(16) and Champs(48) are quite successful in obtaining full sequence coverage of highly purified antibodies. To determine the antibody sequences *de novo*, all these software tools require large number of overlapping peptides, spanning the entire sequence, which are successfully fragmented and converted into predicted peptide sequence reads (**Figure 2a**). This necessitates generation of BU MS data by using multiple proteases. While complicating sample preparation and increasing the required amount, such multi-protease approaches are advantageous for *de novo* sequencing by alleviating the sequence assembly problem.

Chapter 1

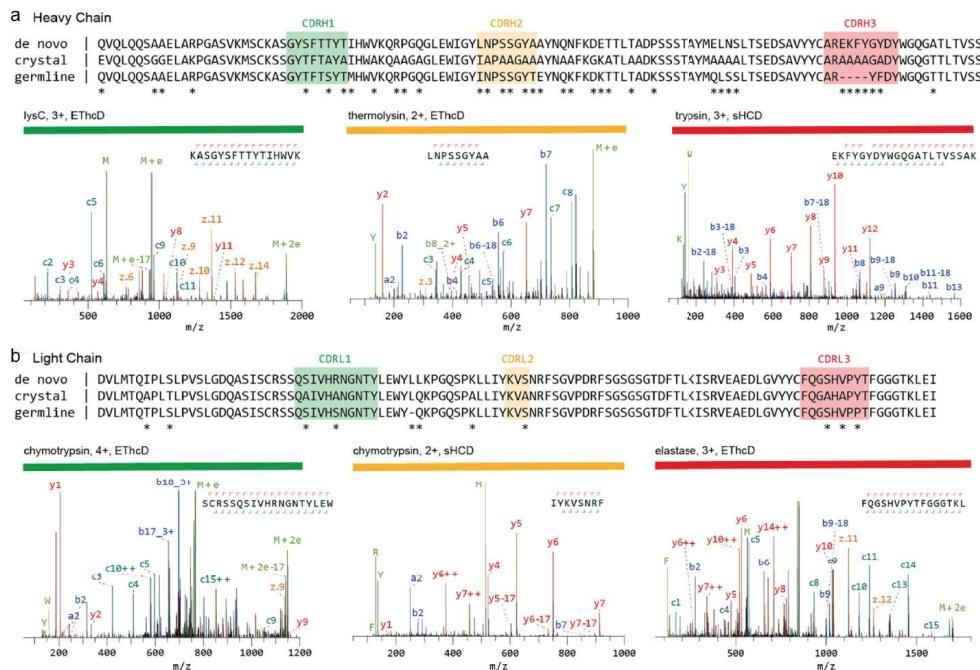


Figure 3: Sequencing of a monoclonal Anti-FLAG M2 antibody. The variable regions of the heavy (a) and light chains (b) are shown. The de novo sequence derived by MS is shown on top, alongside the previously published sequence in the crystal structure of the Fab (PDB ID: 2G60), and germline sequence (IMGT-DomainGapAlign; IGHV1-04/IGHJ2; IGKV1-11*7/IGKJ1). Differential residues are highlighted by asterisks (*). Exemplary MS/MS spectra in support of the assigned sequences are shown below the alignments, labelled with protease, precursor charge state, and fragmentation type. The peptide sequence and fragment coverage are indicated in the top-right of each spectrum spectra, with b/c ions indicated in blue/teal and y/z ions in red/orange. The same colour annotation is used for peaks in the spectra, with additional peaks such as intact/charge reduced precursors, neutral losses, and immonium ions indicated in green. To prevent overlapping peak labels, only a subset of successfully matched peaks are annotated. Figure and caption adapted from Peng et al.(17)

Benefits of complementary peptide fragmentation techniques

In MS-based sequencing, extensive fragmentation of peptide ions is essential to generate arrays of adjacent fragments that reveal the amino acid sequence, often referred to as ion ladders or sequence tags (**Figure 4a**). The amino acid sequence of the fragmented peptide is derived by comparing the mass difference between two adjacent fragment ion peaks to the masses of amino acids and combinations thereof. The produced fragment ion series must contain very few gaps larger than a single amino acid residue, because such gaps lead to exponential growth of the amino acid combinations that fit the mass difference, particularly for spectra of lower

resolution.(49) Since there is no universal fragmentation method that can produce uninterrupted fragment ion ladders for all possible peptides, it is highly advantageous to use various fragmentation methods with distinct mechanisms and specificities to complement each other (**Figure 4b**).⁽⁵⁰⁾ While collisional dissociation (CID/CAD/HCD) is the most used technique in shotgun proteomics experiments, multiple alternative fragmentation techniques have been introduced and have proven to be complementary. These specificities stem from the unique ion activation mechanisms employed by each method. In collision-based techniques, energy is deposited to the multiply protonated peptide ions through low-energetic collisions with inert neutral atoms or gas molecules. This energy is redistributed vibrationally throughout the peptide backbone, fragmenting the most labile bonds and yielding *b/y*-type fragment ions, as defined by the Roepstorff-Fohlmann-Biemann ion nomenclature (**Figure 4c**).⁽⁵¹⁾ Although protonated amide bonds are usually the most susceptible to fragmentation, collisional dissociation often also leads to loss of labile PTMs, such as phosphorylation and sialylation. In electron-based techniques (ECD/ETD), positively-charged peptide ions capture electrons, leading to the generation of odd-electron species that dissociate promptly without significant vibrational redistribution.^(53–55) In contrast to collisional dissociation, this process is not directed towards the most labile bonds, and produces distinctively *c* and *z* fragment ions through the dissociation of N-Cα bond (**Figure 4c**). Similarly, high-energy photon-based activation techniques (UVPD) also cause bond dissociation without substantial energy redistribution. This is enabled by a number of chromophores along the peptide backbone and results in a wide array of co-occurring fragment ion types (*a/x*, *b/y*, *c/z*), depending on the wavelength used.^(52, 56) Highly energetic fragmentation methods can also lead to *w*-type ions, which involve an amino acid side chain dissociation.^(57, 58) In *de novo* sequencing, this may be advantageous since it allows to distinguish between leucine and isoleucine, which are commonly misassigned because they have an identical mass. While having multiple fragment ion types in a single spectrum can compli-

Chapter 1

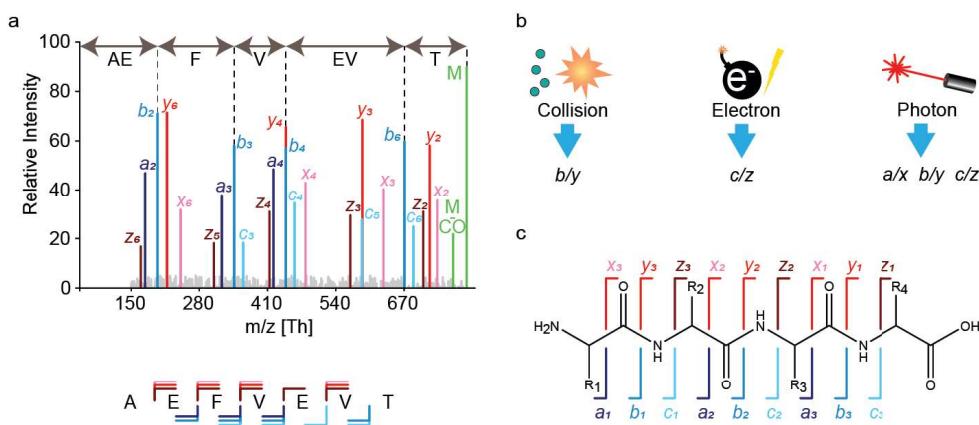


Figure 4: Peptide fragmentation in MS-based de novo sequencing. (a) An illustrative fragmentation spectrum. In the spectrum, fragment ion peaks are colour annotated according to the type of fragment ion (*a*: purple, *b*: blue, *c*: light blue, *x*: pink, *y*: red, and *z*: brown) the unfragmented peptide (precursor ion) is shown in green as well as the precursor ion with neutral loss of CO. Adjacent fragment ions of the same type have a mass difference corresponding to a single amino acid, which is used to determine the sequence as is illustrated for *b*-ions above spectrum. Below the spectrum the amino acid sequence is shown together with the fragment ion annotation, N-terminal fragments (*a*-, *b*-, *c*-) are below the sequence and C-terminal fragments (*x*-, *y*-, *z*-) are shown above the sequence. (b) Three predominant gas phase fragmentation techniques with their predominantly produced fragment ion types. Collisional dissociation (CID/CAD/HCD) predominantly yield *b/y* ions. Electron based dissociation (ECD/ETD) yields *c/z* ions. Contrary the other techniques, high energy photon based dissociation (UVPD) results in all fragment ion types.(52) (c) The Roepstorff-Fohlmann-Biemann nomenclature used for peptide fragment ions denotes different fragment ion types by italic letters *a-c* and *x-z*. The numbering indicates the position of the bond in the amino acid sequence with respect to the N- and C-termini.

cate ion ladder detection (Figure 4a), it can also provide insight into the direction of fragment ion series, revealing to which terminus (N or C) peptide fragments belong. This is possible due to the characteristic mass shift patterns of consecutive *a*, *b*, *c* fragments and consecutive *x*, *y*, *z* fragments originating from the same peptide bond. Horn et al.(59) pioneered this approach for *de novo* protein sequencing by combining CID and ECD to discern between the N- and C-terminal fragment ions, which simplified the detection of consecutive fragment ions. Subsequently, many others have used similar strategies.(60–64) The previously described publication by Peng et al.(17) also demonstrates the successful application of using multiple fragmentation techniques. They recorded spectra using a dual fragmentation scheme of both high-energy collision dissociation (HCD) and electron-transfer high-energy collision dissociation (EThcD), resulting in a reduced number of sequencing errors

when compared to using a single fragmentation method. The spectra selected to support the CDR predictions are also derived from both fragmentation techniques, showing that this versatile fragmentation strategy can benefit sequence coverage in these challenging and important regions (**Figure 3**). Such multiplexing MS strategies have made *de novo* sequencing of mAbs feasible, at least when they are of sufficient purity. However, the procedure is quite laborious as it often involves using multiple proteases to generate overlapping peptides and multiple peptide fragmentation techniques to obtain unambiguous sequence reads, which entails longer sample preparation time, the requirement of larger sample amounts, and extensive data acquisition.

Homology-aided *de novo* sequencing of antibodies

To identify peptides and proteins, shotgun proteomics experiments rely on matching observed fragmentation spectra to theoretical spectra generated from sequence databases. However, complete and accurate mature sequences are not generally available for many proteins, especially for highly variable or frequently mutated proteins like antibodies. Instead, homologous sequences, primarily derived from genomic or transcriptomic experiments, can be used. For antibodies, the genes encoding for each of the regions (V, D, J, and C) are available as germline sequences and can be retrieved from the IMGT database.(25, 65) While such a database of homologous sequences can facilitate verification or guide predictions of *de novo* sequences, it should be noted that the exact match to the target sequence is likely not present even in the most extensive databases. Traditional database searches are thus not applicable because they require exact mass matching of fragments, and a single amino acid mutation can prevent identification. Instead, error-tolerant fragment matching algorithms, either based on sequence alignments or subsequence (i.e., sequence tag) extractions, can use homologous databases to score experimentally determined sequences. An example of a homology-aided approach is searching BU MS data from

a sample of human antibodies against a proteome database such as Swiss-Prot, whereafter the identified peptides are aligned to the IMGT database.(66, 67) Further reported adaptations include *de novo* sequencing of unidentified features from the initial search with dedicated tools, such as PEAKS, to sequence and identify hyper-variable regions.(68, 69) Homology-aided *de novo* sequencing algorithms are also advantageous in identifying erroneous *de novo* peptide reads by comparing them against homologous sequences. In addition, they can be used as a germline template to aid in the assembly of *de novo* peptide reads. Alternative to scaffolds based on homologous sequences, accurate masses of the antibody clones and constituent parts, e.g., light chain or heavy chain, can create mass-based scaffolds. However, these masses need to be obtained separately by performing additional protein-centric MS experiments.

Protein-centric MS approaches

Although conventional *de novo* sequencing of proteins predominantly follows a peptide-centric approach, there have been various attempts to analyse recombinant mAbs intact or at the level of large domains, e.g., Fabs, bringing along a new set of challenges. First, compared to peptides, intact proteins sometimes ionize less efficiently, and liquid-chromatography-based separation of peptides is more established and efficient than separation of intact proteins. Furthermore, in MS analysis, mass accuracy and resolution typically diminish with increasing molecular weight, even when using the latest high-resolution mass spectrometers.(29, 70, 71) In addition, full sequence coverage is generally unattainable for intact proteins with masses above 20 kDa. These factors have significantly held back the implementation of protein-centric MS for *de novo* sequencing of antibodies. However, more recently, several advances in the field resulted in relatively high sequence coverages, reported for recombinant mAbs with available reference sequences.(35, 72–74) Protein-centric approaches, termed TD MS,(32) can provide additional valuable information, including

the mass of the intact antibody,(29) masses of the light and heavy chains, and some predictable fragment ions, which could be used as mass constraints.(18, 30, 75–77) Similar to peptide-centric strategies, there is the potential to combine multiple fragmentation techniques in TD MS to boost sequence coverage. In addition, intact antibody sequencing can be simplified by reducing the complexity and size of the antibody through disulphide reduction or by digestion of the antibodies using specific proteases, such as IgdE (commercially termed FabALACTICA), which cleaves above the hinge region of IgG1, specifically producing 50 kDa Fab fragments,(78) or IdeS (FabRICATOR), a cysteine protease that digests antibodies at a specific site below the hinge, generating F(ab')2 fragments of all IgG subclasses.(34) Such strategies deviate from intact protein sequencing, which resulted in the introduction of the term MD MS.(79) However, these MD strategies still adhere to the core principles of protein-centric MS, whereby large (50-100 kDa) domains of antibodies are analysed. In a large body of works, Fornelli et al.(35, 80–82) have shown how various factors, including sample preparation strategies, fragmentation conditions, and other improvements in instrumentation and experimental design, influence sequence coverage in the protein-centric analysis of recombinant mAbs. Recently, Shaw et al.(83) demonstrated that with modern instrumentation it is possible to successfully fragment intact mAbs in their native state (**Figure 5**). By combining ECD and HCD in a single tandem MS experiment, 42% sequence coverage for the light chain (**Figure 5a**) and 20% sequence coverage for the heavy chain (**Figure 5b**) of Trastuzumab were obtained. The resulting fragmentation spectrum contained not only the multiply charged backbone fragmentation products but also the intact light chain, ejected from the antibody by fragmentation of the intermolecular disulphide bridge, providing information on the light and heavy chain pairing (**Figure 5c**). These and many other studies culminated in a large joint effort by the Consortium for Top-Down Proteomics, wherein they comprehensively described available approaches, techniques, and instrumentation for the analysis of recombinant mAbs.(18) Electron-based fragmentation of

intact protein ions holds great potential for mAb sequencing. Several recent studies showed that these methods consistently yielded nearly uninterrupted c-ion ladders spanning the CDR3, which is paramount to antigen binding.(30, 75, 76, 83, 84) These studies also demonstrated for various antibody isotypes (IgG1-4 and IgA1) that electron-based fragmentation methods consistently provide fragments containing the entire variable region of both the light and heavy chain. Notably, very similar fragments were formed for the intact mAb, the F(ab')2 (produced with IdeS enzyme), and Fab molecules (produced with IgdE or Operator enzymes), showing that reducing antibody complexity through the removal of the Fc portion is not detrimental for protein-centric analysis of mAbs. While significant advances have been made in protein-centric sequencing of purified recombinant antibodies, studying endogenous antibodies remains much more challenging. The separation of intact proteins by liquid chromatography is typically less efficient than the separation strategies available for peptides.(85) This problem is exacerbated for antibody mixtures since different antibody clones are very similar and only vary in a small fraction of the overall sequence. Such minute differences are easily resolved on the peptide yet are significantly more difficult to distinguish on the level of intact antibodies with more than 1000 amino acid residues. Notwithstanding the challenges of intact protein MS, the prospects and potential benefits that protein-centric approaches bring to the *de novo* analytical toolbox are hard to neglect. While it is still nearly impossible to fully *de novo* sequence intact mAbs, protein-centric sequencing can be combined with peptide-centric methods in a hybrid MS approach, providing complementary information substantially advancing towards the goal of complete antibody sequencing by MS, as further described in the section “Combining peptide- and protein-centric MS approaches for antibody sequencing” (**Figure 2b**).

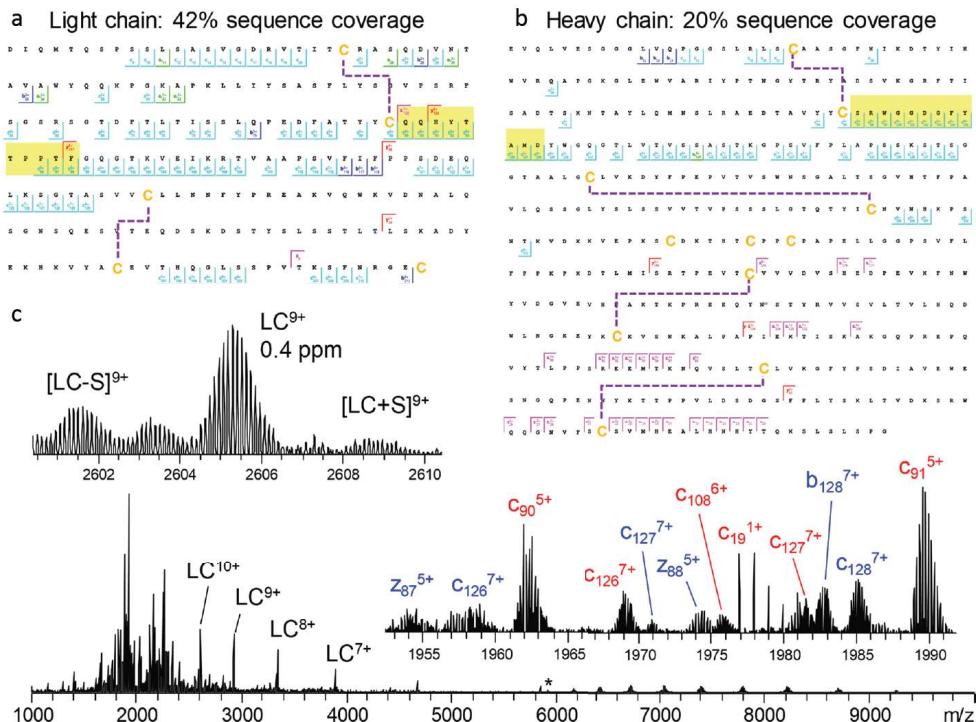


Figure 5: Light chain (a) and glycosylated heavy chain (b) fragmentation maps illustrate sequence coverage produced by the combination of ECD and HCD on Trastuzumab. Disulphide bonds are shown by dashed lines, CDR3 regions are highlighted in yellow. The corresponding fragmentation spectrum (c) for the 25+ charge state of intact Trastuzumab with insets displaying the zoomed in region containing the 9+ charge state of the light chain and various fragment ions. Red and blue fragment ion labels correspond to the light and heavy chain, respectively. Asterisk indicates the mass-selected precursor ion. Figure adapted from Shaw et al.(83)

Dedicated software solutions for MS-based antibody sequencing

The various sample preparation methods and intricate experimental designs presented above result in extended datasets that are not feasible for manual interpretation. Thus, development of dedicated software tools for data interpretation is essential. With regards to BU MS data, presently, two popular software suites are tailored towards *de novo* sequencing of antibodies, SuperNovo(16) and PeaksAB.(27, 86) These suites can utilize the benefits of data generated by using multiple enzymes, multiple fragmentation methods, and the use of a homologous antibody germline sequence database like IMGT to make a complete *de novo* sequence prediction based on the BU MS data. More specifically, the software iteratively screens predicted pep-

Chapter 1

tides against the germline gene segments of the antibody to determine the positions on the final chain construct. Homologous germline sequence candidates represent scaffolds that are then modified to account for the highest scoring predicted peptides. This allows for predicting both heavy and light chain sequences with a minimal error rate of only a few single amino acids per sequenced antibody. A downside, however, is that the software works so far exclusively for sequencing single, highly purified antibodies. Novel software solutions for *de novo* antibody sequencing are emerging and advancing in parallel with improvements in experimental design and instrumentation. The fast development of new *de novo* sequencing strategies encourages the development of new software solutions and improvement of already established tools and requires adaptable software to accommodate the frequent and considerable shifts in *de novo* sequencing approaches, such as the inclusion of TD or MD MS data, multiple fragmentation methods or the analysis of polyclonal samples as opposed to mAbs.

Combining peptide- and protein-centric MS approaches for antibody sequencing

Recent advances in protein-centric MS have spawned various software tools that use these data either in a standalone manner, such as in Twister,(61, 87) or integrate them with BU MS data, as in TBNovo.(88) Twister applies methods similar to those used for BU MS sequencing, recombining individual sequence tags (rather than peptide reads) into longer sequences using a specific implementation of de Bruijn graphs (T-Bruijn graphs) and sequence tag convolution.(61, 87) TBNovo uses sequence tags and precursor masses from TD MS to provide a scaffold for positioning the *de novo* predicted peptide reads to fill the complete sequence. Their analysis makes use of external BU *de novo* sequencing software, PEAKS,(86) and was tested on protein mixtures. TBNovo has not achieved widespread adoption, perhaps due to the software's complexity and because protein-centric MS was still barely practiced at the time of its first release. Although antibody sequencing at the protein level is still not

trivial, it is being applied on a steadily increasing scale in academia and industry. Efforts to extend the sequencing of antibodies to polyclonal mixtures have however proven extremely challenging. The first obstacle is sample availability. While recombinant mAb samples are typically available in milligram quantities, polyclonal antibody samples are often derived from clinical samples and thus will only be available in limited quantities. Because the median concentration of individual clones in plasma is $\sim 1 \mu\text{g/mL}$ the available protein per individual clone is generally orders of magnitude less compared to mAbs.(30) Furthermore, isolation of individual clones is extremely challenging, further complicating the sequencing process as most software tools are exclusively designed for assembling a single antibody and therefore fail when data represents several alike Ig sequences. Additionally, in complex endogenous polyclonal antibody mixtures, key sequence evidence on the hypervariable regions is often not detected due to a dilution effect, whereby sequence information from the conserved regions becomes amplified (as the latter is present in every clone) and thus suppresses the signal of the CDRs, which are unique for all clones. Even though the algorithms developed for mAb sequencing are not directly applicable for polyclonal antibody sequencing, they provide a great starting point for developing new tools.

1.2.2 Hybrid and multi-omics approaches for studying antibody repertoires

One way to further bridge the gap between sequencing of a single purified antibody and those present in bodily fluids, e.g., serum, is to use hybrid or multi-omics strategies. Using a multi-omics approach, for instance, by supplementing BU MS data with genomics or transcriptomics data derived from the same donor, allows bypassing some challenging aspects of genuine *de novo* sequencing, albeit at the cost of a more complex, labour- and data-intensive workflow (**Figure 2c**). Presently, direct *de novo* sequencing of antibodies from a complex mixture is still a tremendous chal-

Chapter 1

lenge. However, integrating complementary information from multiple sources makes it possible to derive valuable data, even on endogenous antibody repertoires. Several approaches have been pioneered recently, as depicted in **Figure 6** and described in more detail below.

Ig-seq

Since the CDRs of the antibodies largely determine antigen specificity, it comes as no surprise that methods specifically targeting CDR-derived peptides have emerged. Notably, the Ig-seq method pioneered by Lavinder et al.(89) in the Georgiou lab applies B-cell sequencing of a given donor to construct a database of putative CDR3 heavy chain peptides. This database is then used to identify and quantify antibodies using CDR-specific tryptic peptides, effectively side-stepping the need for complete *de novo* sequencing (**Figure 6a**). This workflow is very effective because trypsin-targeted residues (arginine and lysine) are found to precede the CDR3 specifically and are found in the relatively conserved FR4 of the heavy chain, ensuring that tryptic peptides contain the heavy chain CDR3 in the majority (>92%) of IgG clones.(90) BU MS is highly optimized for measuring and detecting tryptic peptides, which makes this approach highly effective, as shown when this method was applied to the longitudinal monitoring of influenza antibodies over multiple years. Monitoring the effects of influenza vaccinations showed that ~60% of the response to vaccination originated from pre-existing clonotypes and highlighted the existence and relatively high abundance of broadly protective, non-neutralizing antibodies.(91) Years later, follow-up studies showed that persistent antibodies account for >70% of the serum response over five years, further promoting the efficiency and strength of the Ig-seq method.(92) It should be noted that relying solely on sequences obtained from PBMCs may provide an incomplete database,(31) as it is only feasible to obtain a subset of PBMCs for analysis. Nonetheless, Ig-seq presents one of the most efficient and successful approaches to analyse and identify clones in Ig repertoires and

monitor how they (dis)appear following a change in physiology, e.g., infection or vaccination.

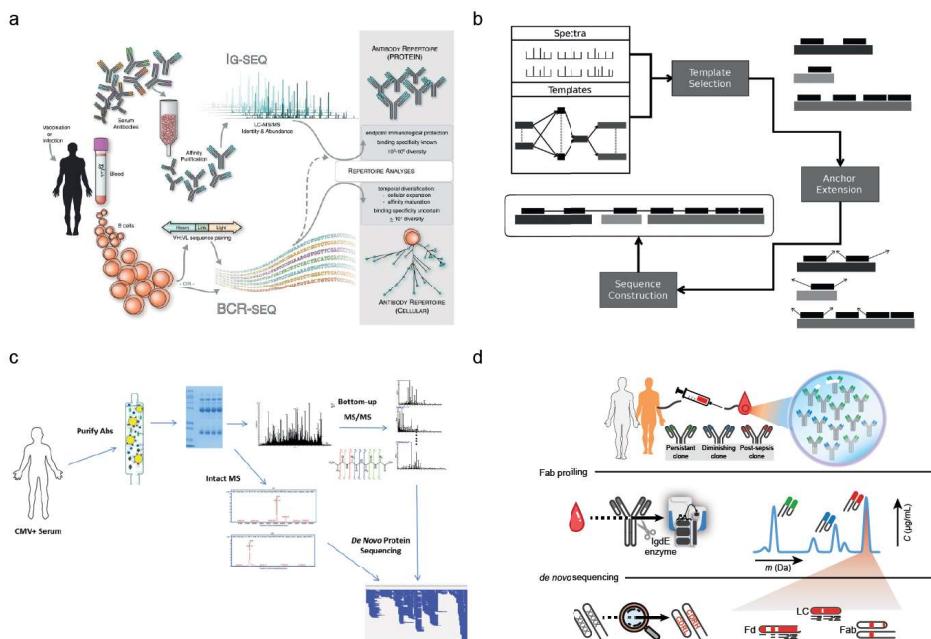


Figure 6: Selected recent approaches aiming towards MS-based de novo sequencing of serum antibodies. (a) In Ig-seq(89) a personalized database generated by BCR sequences is used to identify specific clones, using tryptic peptides covering the CDR3 region. Figure adapted from Lavinder et al.(89) (b) Template proteogenomics(46) use genomic data to generate template sequences. The specific construction of the templates can be defined by the user from either whole genome sequencing or BCR sequencing data. Figure adapted from Castellana et al.(46) (c) PolyExtend(31) helped to analyse a polyclonal mixture of antigen-specific purified antibodies measured by BU MS and intact mass measurements. Using a user-assisted algorithm, these data from different MS modalities were combined to sequence the most abundant clones. Figure adapted from Guthals et al.(31) (d) Fab profiling(30) measures and quantifies intact masses of Fabs to provide a view of the IgG1 clonal repertoire, enabling to quantify and monitor individual clones. Abundant serum clones are identified by using BU and MD MS data iteratively to generate full IgG de novo sequences. Figure adapted from Bondt et al.(30)

Alternative proteogenomics approaches

Extending beyond the Ig-seq strategy, proteogenomics approaches as taken by Castellana et al.(46) incorporate personalized genomics data into the antibody sequencing workflow to identify complete antibody sequences. In their software package GenoMS(46) they accept both proteomic and genomic databases as input, which

Chapter 1

are used to reconstruct antibody (sub)sequences from BU MS data. The database is used to find a homologous template sequence, whereby missing, mutated, and spliced genes are considered. The software also allows for a high degree of flexibility through user-defined constraints. In addition, users can define how the template database is used, excluding certain genes, or using multiple gene segments (V, D, J, or C) to make up a single sequence (**Figure 6b**). As often occurs with hybrid approaches, the power of this proteogenomics strategy comes at a cost. While broadly applicable and very powerful, the required expertise increases because of the use and combination of multiple omics techniques. However, when successfully applied, this workflow produces exciting results, as recently shown in the analyses of antibodies from immunized rabbits(93) and the characterization of neutralizing antibodies against the Ebola virus antigen(94) with notable improvements in integration and visualization of the data. Unfortunately, not all these tools are currently publicly available, although several underlying protocols are open source.(95–97)

1.2.3 Protein-centric sequencing of endogenous antibodies

Some attempts have emerged aiming at novel antibody discovery by MS-based sequencing alone, directly from serum samples or other liquid biopsies, circumventing the need for genomics/transcriptomics data (multi-omics approaches). Above we reviewed several techniques for sequencing purified antibodies. As we pointed out, these methods are geared towards highly purified mAb samples and are therefore not directly applicable for polyclonal antibody mixtures. However, advancements in sample preparation, instrumentation, and bioinformatics make it possible to obtain partial and sometimes complete *de novo* sequences of endogenous antibody clones by combining different mass spectrometric techniques, as discussed further below.

Antigen-specific capture

For many pathologies, it is common to screen patient's serum for antibodies that exhibit activity against the antigens originating from the pathogen, for example, by enzyme-linked immunosorbent assay (ELISA). Using pathogen-based antigens, it is also possible to capture specific antibody clones from serum that exhibit high affinity against the antigen. This typically reduces the complexity of the antibody mixture substantially. Nevertheless, it is still nearly impossible to reduce the complexity down to a single clone, as often, multiple antibodies with varying affinities for any given antigen co-occur. An example of a capturing method whereby additional intact mass data was used to derive *de novo* sequences was described by Guthals et al.(31) Following affinity purification of antibodies from the serum of a cytomegalovirus-exposed individual, using the glycoprotein B antigen, both intact mass and BU MS measurements were performed. Their semi-automated software PolyExtend seeks to use the intact mass measurements to retrieve the average mass of the most abundant species in an antibody mixture, which in turn is used as a mass constraint for a sequence derived using the BU MS data (**Figure 6c**). PolyExtend builds further upon the meta-SPS algorithm,(28) which was initially designed to extend subsequences by assembling multiple sequence predictions into longer subsequences. However, diverging extensions for the same subsequence are treated as sequencing errors with one extension selected for the output. In the case of antibodies, such divergences may indicate the presence of two similar clones. To account for this, the software displays the possible extensions as a ranked list, and the user can then select the extension. This approach aims at expanding the *de novo* sequencing capabilities of the previously established meta-SPS algorithm to deal with simultaneous presence of multiple clones, and Guthals et al.(31) demonstrated a clear proof of concept.

Antibody profiling and sequencing in polyclonal mixtures

While it is still not possible to *de novo* sequence entire serum antibody repertoires, recent advances in LC-MS of intact proteins enabled detecting and resolving single clones from complex antibody mixtures. For instance, developments have been made that specifically profile intact light chains from serum, even providing partial sequence information by using MD MS. Impressively these studies successfully demonstrate the analysis in serum without requiring antigen-specific capture, although they used either a spiked-in mAb as a model or worked with disease models that cause monoclonal Ig overexpression in serum (mono-gammopathy) such as multiple myeloma. Nonetheless, these studies demonstrated that detection and characterization of individual endogenous light chains is possible.(98–102) Taking this one step further, Wang et al.(103) developed a method to detect individual Fab fragments in serum. They were able to identify tens of heavy and light chains of serum autoantibodies. Although attempts were made to *de novo* sequence these antibodies at the intact protein level, the obtained results were limited to a few sequence tags. In light of the SARS-CoV-2 pandemic, Melani et al.(104) focused their profiling efforts on the vaccine-targeted spike protein receptor-binding domain. The approach is named Ig-MS and features two novel metrics for capturing the intensity and complexity of the antibody response. In short, the method uses affinity purification to capture antigen-specific clones. A mAb-containing standard is spiked in for quantitation, and the sample is disulphide-reduced. After reduction, individual ion MS(105) is used to measure a mass fingerprint of the sample. The ratio between the intensity of clonal peaks and the standard is used to estimate the response (“Ion Titer”), and the complexity of the response (“Degree of Clonality”) is assessed by the ratio of the most intense light chain peak to that of the summed intensity of all light chain peaks. Finally, these metrics are correlated to the ELISA-based antibody titer and neutralization efficiency, to verify their accuracy. In a recent study, Bondt et al.(30) used an approach to generate

Fab fragments exclusively from the entire IgG1 repertoire. They were able to longitudinally profile IgG1 Fabs from the serum of both healthy and sepsis-inflicted donors without the need for any enrichment of specific clones. They observed a range of 50-500 distinct detectable IgG1 Fab clones per donor and showed that most clones persist over multiple months of sampling. Contrary to widely held belief, they showed that the IgG1 repertoire is in abundance dominated by just a few hundred clones and that each donor exhibits a unique repertoire of clones. They also managed to directly *de novo* sequence a single highly abundant clone in one of the donors without the aid of antigen-specific capture. The *de novo* sequencing was achieved by a combination of protein-centric sequencing using ETD, and a BU MS approach using multiple proteases for digestion. First, closely matching light and heavy chain germline templates were selected from the IMGT database. Subsequently, the data was used to refine these templates iteratively, yielding the final mature sequence. This provided proof of concept that *de novo* sequencing of clones directly from serum is feasible, although still arduous and limited to specific cases (**Figure 6d**). Notably, the determined sequences contained more mutations (compared to germline sequences) than expected from the reported rates from BCR sequencing studies,(106) which is indicative of potential discrepancies between protein-level and gene-level sequencing. This first attempt focused exclusively on IgG1, by using an IgG1-specific protease to generate the Fab fragments. In another work, Bondt et al.(107) extended their method to IgA1 by using a protease specific to the O-glycans present in IgA1 hinge region to generate Fab fragments, albeit now exclusively from IgA1. Overall, they showed that – similar to serum IgG1 – just a handful of clones dominates the secretory IgA1 profile of human milk. Using a somewhat comparable approach, Dupré et al.(102) analysed isolated light chains from the urine of a patient affected by multiple myeloma. They assembled *de novo* data from peptides into a full-length sequence, using the intact mass data as a scaffold. Subsequently, they used TD MS to validate their findings and further characterize the proteoforms of the light chains, including PTMs. The BU

MS data further supported the resulting proteoforms, showing a similar added benefit from iteratively combining BU and TD MS data.

1.2.4 Additional benefits of studying antibodies at the protein level

The capabilities of MS allow for antibody characterization beyond the primary amino acid sequence. Antibodies are known to harbour multiple important PTMs: Fab- and Fc-glycosylation,(108) deamidation,(109) and C-terminal truncation,(110) to name a few. Moreover, although the disulphide bonds in IgG1 are thoroughly described, other subclasses, notably IgG2, appear to occur as structural isomers induced by different disulphide-bridge patterns. These PTMs and disulphide bridges become even more pronounced in IgA and IgM, which can form higher-order structures connected by the joining-chain in serum and other bodily fluids. All these features influence the antibody's efficacy and stability. Such information cannot easily be obtained at the nucleotide level, requiring protein-level analysis.

1.3 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 subsequent chapters and an important example of the central theme.

Chapter 2 describes how we developed CrossID, a tool to analyze large and complex cross-linking proteomic datasets. CrossID was developed to facilitate explorative 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 cohort of sepsis patients we found the serological IgG1 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 novo* sequencing of endogenous antibodies by using a multi-tier mass spectrometry approach to sequence the most abundant clone for a donor.

Chapter 4 describes the analysis of breastmilk SIgA1 profiles of six mothers who had received two identical SARS-CoV-2 vaccinations over 16 timepoints. We use

Chapter 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 novo* sequencing of endogenous antibodies by hybrid mass spectrometry. We present a more standardized workflow for sequencing antibody chains in mixtures. Our approach resolves ambiguity in sequence predictions for the hypervariable complementarity determining 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.

References

- [1] E S London. Der gegenwärtige stand der lehre von den cytolyssinen und die cytotlytische theorie der immunität. *Zentralbl. f. Bakteriol., Parasitenk. u. Infektionskrankh., Abt. II, Bd. 32*, 1902.
- [2] Jean Lindenmann. Origin of the terms 'antibody' and 'antigen'. *Scandinavian Journal of Immunology*, 19(4):281–285, 1984. ISSN 0300-9475. doi: 10.1111/j.1365-3083.1984.tb00931.x.
- [3] Claire Marks and Charlotte M Deane. How repertoire data are changing antibody science. *The Journal of biological chemistry*, 295(29):9823–9837, 2020. ISSN 1083-351. doi: 10.1074/jbc.rev120.010181.
- [4] Matthew I J Raybould, Claire Marks, Alan P Lewis, et al. Thera-sabdb: the therapeutic structural antibody database. *Nucleic acids research*, 48(1):383–388, 2020. ISSN 1362-4962. doi: 10.1093/nar/gkz827.
- [5] Hélène Kaplon and Janice M. Reichert. Antibodies to watch in 2021. *mAbs*, 13(1), 2021. ISSN 19420870. doi: 10.1080/19420862.2020.1860476.
- [6] Surjit Singh, Nitish K. Kumar, Pradeep Dwiwedi, et al. Monoclonal antibodies: A review. *Current Clinical Pharmacology*, 13(2):85–99, 2018. ISSN 15748847. doi: 10.2174/1574884712666170809124728.
- [7] Harry W. Schroeder and Lisa Cavacini. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*, 125(2):41–52, 2010. ISSN 00916749. doi: 10.1016/j.jaci.2009.09.046.
- [8] Zachary A. Bornholdt, Hannah L. Turner, Charles D. Murin, et al. Isolation of potent neutralizing antibodies from a survivor of the 2014 ebola virus outbreak. *Science*, 351(6277):1078–1083, 2016. ISSN 1095-9203. doi: 10.1126/science.aad5788.
- [9] Davide Corti, John Misasi, Sabue Mulangu, et al. Protective monotherapy against lethal ebola virus infection by a potently neutralizing antibody. *Science (New York, N.Y.)*, 351(6279):1339–1342, 2016. ISSN 1095-9203. doi: 10.1126/science.aad5224.
- [10] Rut Valgardsdottir, Irene Cattaneo, Gavino Napolitano, et al. Identification of human sars-cov-2 monoclonal antibodies from convalescent patients using ebv immortalization. *Antibodies*, 10(3):26, 2021. ISSN 2073-4468. doi: 10.3390/antib10030026.
- [11] H W Schroeder Jr. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev Comp Immunol*, 30(12):119–135, 2006. doi: 10.1016/j.dci.2005.06.006.
- [12] Bryan Briney, Anne Inderbitzin, Collin Joyce, and Dennis R. Burton. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature*, 566(7744):393–397, 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-0879-y.
- [13] Jennifer R. Hom, Deepak Tomar, and Christopher M. Tipton. Exploring the diversity of the b-cell receptor repertoire through high-throughput sequencing. In *Methods in Molecular Biology: Immune Receptors*, pages 231–241. Clifton (NJ), 2022.
- [14] A. F. Maarten Altelaar, Javier Munoz, and Albert J. R. Heck. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics*, 14(1):35–48, 2013. ISSN 1471-0056. doi: 10.1038/nrg3356.
- [15] Ruedi Aebersold and Matthias Mann. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355, 2016. ISSN 0028-0836. doi: 10.1038/nature19949.
- [16] K. Ilker Sen, Wilfred H Tang, Shruti Nayak, et al. Automated antibody de novo sequencing and its utility in biopharmaceutical discovery. *Journal of the American Society for Mass Spectrometry*, 28(5):803–810, 2017. ISSN 1879-1123. doi: 10.1007/s13361-016-1580-0.
- [17] Weiwei Peng, Matti F. Pronker, and Joost Snijder. Mass spectrometry-based de novo sequencing of monoclonal antibodies using multiple proteases and a dual fragmentation scheme. *Journal of Proteome Research*, 20(7):3559–3566, 2021. ISSN 1535-3893. doi: 10.1021/acs.jproteome.1c00169.

Chapter 1

- [18] Kristina Srzentić, Luca Fornelli, Yury O. Tsybin, et al. Inter-laboratory study for characterizing monoclonal antibodies by top-down and middle-down mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 31(9):1783–1802, 2020. ISSN 1044-0305. doi: 10.1021/jasms.0c00036.
- [19] Mark L. Chiu, Dennis R. Goulet, Alexey Teplyakov, and Gary L. Gilliland. Antibody structure and function: The basis for engineering therapeutics. *Antibodies (Basel, Switzerland)*, 8(4):55, 2019. ISSN 2073-4468. doi: 10.3390/antib8040055.
- [20] R R Porter. The hydrolysis of rabbit γ -globulin and antibodies with crystalline papain. *Biochemical Journal*, 73(1):119–127, 1959. ISSN 0306-3283. doi: 10.1042/bj0730119.
- [21] Andrew B. Herr, Edward R. Ballister, and Pamela J. Bjorkman. Insights into iga-mediated immune responses from the crystal structures of human fcalphari and its complex with iga1-fc. *Nature*, 423(6940):614–20, 2003. ISSN 0028-0836. doi: 10.1038/nature01685.
- [22] Jr Charles A Janeway, Paul Travers, Mark Walport, and Mark J Shlomchik. The generation of diversity in immunoglobulins. In *Immunobiology: The Immune System in Health and Disease*. New York (NY), 5 edition, 2001. ISBN 0-8153-3642-X.
- [23] Bruce Alberts, Alexander Johnson, Julian Lewis, et al. The generation of antibody diversity. In *Molecular Biology of the Cell*. New York (NY), 4 edition, 2002. ISBN 0-8153-3218-1.
- [24] D J Jeske, J Jarvis, C Milstein, and J D Capra. Junctional diversity is essential to antibody activity. *Journal of immunology*, 133(3):1090–1092, 1984. ISSN 0022-1767.
- [25] Marie Paule Lefranc. Imgt, the international immunogenetics database®. *Nucleic Acids Research*, 31(1):307–310, 2003. ISSN 03051048. doi: 10.1093/nar/gkg085.
- [26] Ruedi Aebersold and Matthias Mann. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, 2003. ISSN 0028-0836. doi: 10.1038/nature01511.
- [27] Ngoc Hieu Tran, M Ziaur Rahman, Lin He, et al. Complete de novo assembly of monoclonal antibody sequences. *Scientific Reports*, 6(1):31730, 2016. ISSN 2045-2322. doi: 10.1038/srep31730.
- [28] Adrian Guthals, Karl R Clauser, and Nuno Bandeira. Shotgun protein sequencing with meta-contig assembly. *Molecular and Cellular Proteomics*, 11(10):1084–1096, 2012. ISSN 15359476. doi: 10.1074/mcp.m111.015768.
- [29] Daniel P. Donnelly, Catherine M. Rawlins, Caroline J. De-Hart, et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature Methods* 2019 16:7, 16(7):587–594, 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0457-0.
- [30] Albert Bondt, Max Hoek, Sem Tamara, et al. Human plasma igg1 repertoires are simple, unique, and dynamic. *Cell Systems*, 12(12):1131–11435, 2021. ISSN 24054720. doi: 10.1016/j.cels.2021.08.008.
- [31] Adrian Guthals, Yutian Gan, Laura Murray, et al. De novo ms/ms sequencing of native human antibodies. *Journal of Proteome Research*, 16(1):45–54, 2017. ISSN 1535-3893. doi: 10.1021/acs.jproteome.6b00608.
- [32] Timothy K. Toby, Luca Fornelli, and Neil L. Kelleher. Progress in top-down proteomics and the analysis of proteoforms. *Annual Review of Analytical Chemistry*, 9(1):499–519, 2016. ISSN 1936-1327. doi: 10.1146/annurev-anchem-071015-041550.
- [33] Philip D Compton, Leonid Zamdborg, Paul M Thomas, and Neil L Kelleher. On the scalability and requirements of whole protein mass spectrometry. *Analytical chemistry*, 83(17):6868–6874, 2011. ISSN 1520-6882. doi: 10.1021/ac2010795.
- [34] Björn P. Johansson, Oonagh Shannon, and Lars Björck. Ides: A bacterial proteolytic enzyme with therapeutic potential. *PLoS ONE*, 3(2):1692, 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0001692.
- [35] Luca Fornelli, Daniel Ayoub, Konstantin Aizikov, Alain Beck, and Yury O Tsybin. Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap fourier transform mass spectrometry. *Analytical Chemistry*, 86(6):3005–3012, 2014. ISSN 0003-2700. doi: 10.1021/ac4036857.
- [36] Robert M Hudziak, Gail D Lewis, Marcy Winget, et al. p185her2 monoclonal antibody has antiproliferative effects in vitro and sensitizes human breast tumor cells to tumor necrosis factor. *Molecular and Cellular Biology*, 9(3):1165–1172, 1989. ISSN 0270-7306. doi: 10.1128/mcb.9.3.1165-1172.1989.

- [37] Paul Carter, L. Presta, Cornelia M Gorman, et al. Humanization of an anti-p185her2 antibody for human cancer therapy. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10):4285–4289, 1992. ISSN 0027-8424. doi: 10.1073/pnas.89.10.4285.
- [38] Naoya Tsurushita, Paul R. Hinton, and Shankar Kumar. Design of humanized antibodies: From anti-tac to zenapax. *Methods*, 36(1):69–83, 2005. ISSN 10462023. doi: 10.1016/j.ymeth.2005.01.007.
- [39] Leila Khoja, Marcus O Butler, S Peter Kang, Scott Ebbinghaus, and Anthony M Joshua. Pembrolizumab. *Journal for ImmunoTherapy of Cancer*, 3(1):36, 2015. ISSN 2051-1426. doi: 10.1186/s40425-015-0078-9.
- [40] William Busse, Jonathan Corren, Bobby Quentin Lanier, et al. Omalizumab, anti-ige recombinant humanized monoclonal antibody, for the treatment of severe allergic asthma. *Journal of Allergy and Clinical Immunology*, 108(2):184–190, 2001. ISSN 00916749. doi: 10.1067/mai.2001.117880.
- [41] L G Presta, S J Lahr, R L Shields, et al. Humanization of an antibody directed against ige. *Journal of immunology*, 151(5):2623–2632, 1993. ISSN 0022-1767.
- [42] Ruei-Min Lu, Yu-Chyi Hwang, I-Ju Liu, et al. Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, 27(1):1, 2020. ISSN 1423-0127. doi: 10.1186/s12929-019-0592-z.
- [43] George Georgiou, Gregory C. Ippolito, John Beausang, et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology*, 32(2):158–168, 2014. ISSN 10870156. doi: 10.1038/nbt.2782.
- [44] Denis Morsa, Dominique Baiwir, Raphaël La Rocca, et al. Multi-enzymatic limited digestion: The next-generation sequencing for proteomics? *Journal of Proteome Research*, 18(6):2501–2513, 2019. ISSN 1535-3893. doi: 10.1021/acs.jproteome.9b00044.
- [45] Nuno Bandeira, Victoria Pham, Pavel Pevzner, David Arnott, and Jennie R Lill. Beyond edman degradation: Automated de novo protein sequencing of monoclonal antibodies. *Nature Biotechnology*, 26(12):1336–1338, 2008. ISSN 1087-0156. doi: 10.1038/nbt1208-1336.
- [46] Natalie E Castellana, Victoria Pham, David Arnott, Jennie R Lill, and Vineet Bafna. Template proteogenomics: Sequencing whole proteins using an imperfect database. *Molecular and Cellular Proteomics*, 9(6):1260–1270, 2010. ISSN 15359476. doi: 10.1074/mcp.m900504-mcp200.
- [47] Alon Savidor, Rotem Barzilay, Dalia Elinger, et al. Database-independent protein sequencing (dips) enables full-length de novo protein and antibody sequence determination. *Molecular and cellular proteomics : MCP*, 16(6):1151–1161, 2017. ISSN 1535-9484. doi: 10.1074/mcp.o116.065417.
- [48] Xiaowen Liu, Yonghua Han, Denis Yuen, and Bin Ma. Automated protein (re)sequencing with ms/ms and a homologous database yields almost full coverage and accuracy. *Bioinformatics*, 25(17):2174–2180, 2009. ISSN 1460-2059. doi: 10.1093/bioinformatics/btp366.
- [49] Lidong He, Chad R. Weisbrod, and Alan G. Marshall. Protein de novo sequencing by top-down and middle-down ms/ms: Limitations imposed by mass measurement accuracy and gaps in sequence coverage. *International Journal of Mass Spectrometry*, 427:107–113, 2018. ISSN 13873806. doi: 10.1016/j.ijms.2017.11.012.
- [50] Luis A. Macias, Inês C. Santos, and Jennifer S. Brodbelt. Ion activation methods for peptides and proteins. *Analytical Chemistry*, 92(1):227–251, 2020. ISSN 0003-2700. doi: 10.1021/acs.analchem.9b04859.
- [51] P. Roepstorff and J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry*, 11(11):601, 1984. ISSN 0306-042. doi: 10.1002/bms.1200111109.
- [52] Jennifer S. Brodbelt. Ion activation methods for peptides and proteins. *Analytical Chemistry*, 88(1):30–51, 2016. ISSN 15206882. doi: 10.1021/acs.analchem.5b04563.
- [53] John E. P. Syka, Joshua J. Coon, Melanie J. Schroeder, Jeffrey Shabanowitz, and Donald F. Hunt. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences*, 101(26):9528–9533, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0402700101.

- [54] Roman A. Zubarev, David M. Horn, Einar K. Fridriksson, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Analytical Chemistry*, 72(3): 563–573, 2000. ISSN 0003-2700. doi: 10.1021/ac990811p.
- [55] Scott A. McLuckey and James L. Stephenson. Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrometry Reviews*, 17(6):369–407, 1998. ISSN 02777037. doi: 10.1002/(sici)1098-2787(1998)17:6<369::aid-mas1>3.0.co;2-j.
- [56] Jennifer S. Brodbelt, Lindsay J. Morrison, and Inês Santos. Ultraviolet photodissociation mass spectrometry for analysis of biological molecules. *Chemical Reviews*, 120(7):3328–3380, 2020. ISSN 0009-2665. doi: 10.1021/acs.chemrev.9b00440.
- [57] Yongsheng Xiao, Małgorzata M. Vecchi, and Dingyi Wen. Distinguishing between leucine and isoleucine by integrated LC-ms analysis using an orbitrap fusion mass spectrometer. *Analytical Chemistry*, 88(21):10757–10766, 2016. ISSN 15206882. doi: 10.1021/acs.analchem.6b03409.
- [58] Frank Kjeldsen, Kim F. Haselmann, Esben S. Sørensen, and Roman A. Zubarev. Distinguishing of ile/leu amino acid residues in the pp3 protein by (hot) electron capture dissociation in fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry*, 75(6):1267–1274, 2003. ISSN 0003-2700. doi: 10.1021/ac020422m.
- [59] D M Horn, R A Zubarev, and F W McLafferty. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proceedings of the National Academy of Sciences*, 97(19):10313–10317, 2000. ISSN 0027-8424. doi: 10.1073/pnas.97.19.10313.
- [60] Adrian Guthals, Karl R. Clauser, Ari M. Frank, and Nuno Bandeira. Sequencing-grade de novo analysis of ms/ms triplets (cid/hcd/etd) from overlapping peptides. *Journal of Proteome Research*, 12(6):2846–2857, 2013. ISSN 15353893. doi: 10.1021/pr400173d.
- [61] Kira Vyatkina. De novo sequencing of top-down tandem mass spectra: A next step towards retrieving a complete protein sequence. *Proteomes*, 5(1):6, 2017. ISSN 2227-7382. doi: 10.3390/proteomes5010006.
- [62] Ritendra Datta and Marshall Bern. Spectrum fusion: Using multiple mass spectra for de novo peptide sequencing. *Journal of Computational Biology*, 16(8):1169–1182, 2009. ISSN 1066-5277. doi: 10.1089/cmb.2009.0122.
- [63] Andrew P. Horton, Scott A. Robotham, Joe R. Cannon, et al. Comprehensive de novo peptide sequencing from ms/ms pairs generated through complementary collision induced dissociation and 351 nm ultraviolet photodissociation. *Analytical Chemistry*, 89(6):3747–3753, 2017. ISSN 0003-2700. doi: 10.1021/acs.analchem.7b00130.
- [64] Andreas Bertsch, Andreas Leinenbach, Anton Pervukhin, et al. De novo peptide sequencing by tandem ms using complementary cid and electron transfer dissociation. *Electrophoresis*, 30(21):3736–3747, 2009. ISSN 01730835. doi: 10.1002/elps.200900332.
- [65] Marie-Paule Lefranc and Gérard Lefranc. Immunoglobulins or antibodies: IgM® bridging genes, structures and functions. *Biomedicines*, 8(9):319, 2020. ISSN 2227-9059. doi: 10.3390/biomedicines8090319.
- [66] Carsten Schmelter, Natarajan Perumal, Sebastian Funke, et al. Peptides of the variable IgG domain as potential biomarker candidates in primary open-angle glaucoma (poag). *Human Molecular Genetics*, 26(22):4451–4464, 2017. ISSN 0964-6906. doi: 10.1093/hmg/ddx332.
- [67] Vaibhav Singh, Marcel P. Stoop, Christoph Stingl, et al. Cerebrospinal-fluid-derived immunoglobulin G of different multiple sclerosis patients shares mutated sequences in complementarity determining regions. *Molecular and Cellular Proteomics*, 12(12):3924–3934, 2013. ISSN 15359476. doi: 10.1074/mcp.m113.030346.
- [68] Ingrid Broodman, Dominique de Costa, Christoph Stingl, et al. Mass spectrometry analyses of κ and λ fractions result in increased number of complementarity-determining region identifications. *Proteomics*, 12(2):183–191, 2012. ISSN 16159853. doi: 10.1002/pmic.201100244.
- [69] Dominique de Costa, Ingrid Broodman, Martijn M. VanDuijn, et al. Sequencing and quantifying IgG fragments and antigen-binding regions by mass spectrometry. *Journal of Proteome Research*, 9(6):2937–2945, 2010. ISSN 1535-3893. doi: 10.1021/pr901114w.
- [70] Sem Tamara, Maurits A. den Boer, and Albert J. R. Heck. High-resolution native mass spectrometry. *Chemical reviews*, page 100212, 2021. ISSN 1520-6890. doi: 10.1021/acs.chemrev.1c00212.

- [71] Philip Lössl, Joost Snijder, and Albert J. R. Heck. Boundaries of mass resolution in native mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 25(6):906–917, 2014. ISSN 1044-0305. doi: 10.1007/s13361-014-0874-3.
- [72] Yuan Mao, Santosh G Valeja, Jason C Rouse, Christopher L Hendrickson, and Alan G Marshall. Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-fourier transform ion cyclotron resonance-mass spectrometry. *Analytical chemistry*, 85(9):4239–4246, 2013. ISSN 1520-6882. doi: 10.1021/ac303525n.
- [73] Yury O Tsybin, Luca Fornelli, Carsten Stoermer, et al. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Analytical Chemistry*, 83(23):8919–8927, 2011. ISSN 0003-2700. doi: 10.1021/ac201293m.
- [74] Anja Resemann, Wolfgang Jabs, Anja Wiechmann, et al. Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing. *mAbs*, 8(2):318–330, 2016. ISSN 1942-0862. doi: 10.1080/19420862.2015.1128607.
- [75] Jean-Francois Greisch, Maurits A. den Boer, Frank Beurskens, et al. Generating informative sequence tags from antigen-binding regions of heavily glycosylated IgA1 antibodies by native top-down electron capture dissociation. *Journal of the American Society for Mass Spectrometry*, 32(6):1326–1335, 2021. ISSN 1044-0305. doi: 10.1021/jasms.0c00461.
- [76] Maurits A. den Boer, Jean-Francois Greisch, Sem Tamara, Albert Bondt, and Albert J. R. Heck. Selectivity over coverage in de novo sequencing of IgGs. *Chemical science*, 11(43):11886–11896, 2020. ISSN 2041-6520. doi: 10.1039/d0sc03438j.
- [77] Jean-Francois Greisch, Maurits A. den Boer, Szu-Hsueh Lai, et al. Extending native top-down electron capture dissociation to mAb immunoglobulin complexes provides useful sequence tags covering their critical variable complementarity-determining regions. *Analytical Chemistry*, 93(48):16068–16075, 2021. ISSN 0003-2700. doi: 10.1021/acs.analchem.1c03740.
- [78] Christian Spoerry, Jana Seele, Peter Valentin-Weigand, Christoph G. Baums, and Ulrich von Pawel-Rammingen. Identification and characterization of Igde, a novel IgG-degrading protease of *streptococcus suis* with unique specificity for porcine IgG. *Journal of Biological Chemistry*, 291(15):7915–7925, 2016. ISSN 0021-9258. doi: 10.1074/jbc.m115.711440.
- [79] Frederik Lermyte, Yury O. Tsybin, Peter B. O'Connor, and Joseph A. Loo. Top or middle? up or down? toward a standard lexicon for protein top-down and allied mass spectrometry approaches. *Journal of the American Society for Mass Spectrometry*, 30(7):1149–1157, 2019. ISSN 1044-0305. doi: 10.1007/s13361-019-02201-x.
- [80] Luca Fornelli, Eugen Damoc, Paul M. Thomas, et al. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation orbitrap FTMS. *Molecular and Cellular Proteomics*, 11(12):1758–1767, 2012. ISSN 15359476. doi: 10.1074/mcp.m112.019620.
- [81] Luca Fornelli, Daniel Ayoub, Konstantin Aizikov, et al. Top-down analysis of immunoglobulin G isotypes 1 and 2 with electron transfer dissociation on a high-field orbitrap mass spectrometer. *Journal of Proteomics*, 159:67–76, 2017. ISSN 1876-7737. doi: 10.1016/j.jprot.2017.02.013.
- [82] Luca Fornelli, Kristina Srzentić, Romain Huguet, et al. Accurate sequence analysis of a monoclonal antibody by top-down and middle-down orbitrap mass spectrometry applying multiple ion activation techniques. *Analytical Chemistry*, 90(14):8421–8429, 2018. ISSN 0003-2700. doi: 10.1021/acs.analchem.8b00984.
- [83] Jared B. Shaw, Weijing Liu, V. Yury Vasilev, et al. Direct determination of antibody chain pairing by top-down and middle-down mass spectrometry using electron capture dissociation and ultraviolet photodissociation. *Analytical Chemistry*, 92(1):766–773, 2020. ISSN 1520-6882. doi: 10.1021/acs.analchem.9b03129.
- [84] Jared B. Shaw, Neha Malhan, V. Yury Vasilev, et al. Sequencing grade tandem mass spectrometry for top-down proteomics using hybrid electron capture dissociation methods in a benchtop orbitrap mass spectrometer. *Analytical Chemistry*, 90(18):10819–10827, 2018. ISSN 1520-6882. doi: 10.1021/acs.analchem.8b01901.

Chapter 1

- [85] Yufeng Shen, Nikola Tolić, Paul D Piehowski, et al. High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics. *Journal of chromatography. A*, 1498:99–110, 2017. ISSN 1873-3778. doi: 10.1016/j.chroma.2017.01.008.
- [86] Bin Ma, Kaizhong Zhang, Christopher Hendrie, et al. Peaks: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 17(20):2337–2342, 2003. ISSN 0951-4198. doi: 10.1002/rcm.1196.
- [87] Kira Vyatkina, Si Wu, Lennard J. M. Dekker, et al. De novo sequencing of peptides from top-down tandem mass spectra. *Journal of Proteome Research*, 14(11):4450–4462, 2015. ISSN 1535-3893. doi: 10.1021/pr501244v.
- [88] Xiaowen Liu, Lennard J M Dekker, Si Wu, et al. De novo protein sequencing by combining top-down and bottom-up tandem mass spectra. *Journal of Proteome Research*, 13(7):3241–3248, 2014. ISSN 1535-3893. doi: 10.1021/pr401300m.
- [89] Jason J. Lavinder, Andrew P. Horton, George Georgiou, and Gregory C. Ippolito. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Current opinion in chemical biology*, 24:112–20, 2015. ISSN 1879-0402. doi: 10.1016/j.cbpa.2014.11.007.
- [90] Jason J. Lavinder, Yariv Wine, Claudia Giesecke, et al. Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proceedings of the National Academy of Sciences*, 111(6):2259–2264, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1317793111.
- [91] Jiwon Lee, Daniel R. Boutz, Veronika Chromikova, et al. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nature Medicine*, 22(12):1456–1464, 2016. ISSN 1546170. doi: 10.1038/nm.4224.
- [92] Jiwon Lee, Philipp Paparoditis, Andrew P. Horton, et al. Persistent antibody clonotypes dominate the serum response to influenza over multiple years and repeated vaccinations. *Cell Host and Microbe*, 25(3):367–3765, 2019. ISSN 19313128. doi: 10.1016/j.chom.2019.01.010.
- [93] Stefano R. Bonisone, Thiago Lima, Katherine Harris, et al. Serum proteomics expands on high-affinity antibodies in immunized rabbits than deep b-cell repertoire sequencing alone. *bioRxiv*, page 833871, 2020. ISSN 2692-8205. doi: 10.1101/833871.
- [94] Pavlo Gilchuk, Adrian Guthals, Stefano R. Bonisone, et al. Proteo-genomic analysis identifies two major sites of vulnerability on ebolavirus glycoprotein for neutralizing antibodies in convalescent human plasma. *Frontiers in Immunology*, 12:706757, 2021. ISSN 1664-3224. doi: 10.3389/fimmu.2021.706757.
- [95] Seong Won Cha, Stefano Bonisone, Seungjin Na, Pavel A. Pevzner, and Vineet Bafna. The antibody repertoire of colorectal cancer. *Molecular and Cellular Proteomics*, 16(12):2111–2124, 2017. ISSN 15359476. doi: 10.1074/mcp.ra117.000397.
- [96] Yana Safonova, Stefano Bonisone, Eugene Kurpilyansky, et al. Igrepertoireconstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis. *Bioinformatics*, 31(12):53–61, 2015. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv238.
- [97] Stefano R. Bonisone and Pavel A. Pevzner. Immunoglobulin classification using the colored antibody graph. *Journal of Computational Biology*, 23(6):483–494, 2016. ISSN 1066-5277. doi: 10.1089/cmb.2016.0010.
- [98] Lidong He, Lissa C. Anderson, David R. Barnidge, et al. Analysis of monoclonal antibodies in human serum as a model for clinical monoclonal gammopathy by use of 21 tesla ft-icr top-down and middle-down ms/ms. *Journal of the American Society for Mass Spectrometry*, 28(5):827–838, 2017. ISSN 1044-0305. doi: 10.1007/s13361-017-1602-6.
- [99] Lidong He, Lissa C. Anderson, David R. Barnidge, et al. Classification of plasma cell disorders by 21 tesla fourier transform ion cyclotron resonance top-down and middle-down ms/ms analysis of monoclonal immunoglobulin light chains in human serum. *Analytical Chemistry*, 91(5):3263–3269, 2019. ISSN 0003-2700. doi: 10.1021/acs.analchem.8b03294.
- [100] John R. Mills, David R. Barnidge, and David L. Murray. Detecting monoclonal immunoglobulins in human serum using mass spectrometry. *Methods*, 81:56–65, 2015. ISSN 10462023. doi: 10.1016/j.ymeth.2015.04.020.

- [101] Faye A. Sharpley, Richa Manwani, Shameem Mahmood, et al. A novel mass spectrometry method to identify the serum monoclonal light chain component in systemic light chain amyloidosis. *Blood Cancer Journal*, 9(2):16, 2019. ISSN 2044-5385. doi: 10.1038/s41408-019-0180-1.
- [102] Mathieu Dupré, Magalie Duchateau, Rebecca Sternke-Hoffmann, et al. De novo sequencing of antibody light chain proteoforms from patients with multiple myeloma. *Analytical Chemistry*, 93(30):10627–10634, 2021. ISSN 15206882. doi: 10.1021/acs.analchem.1c01955.
- [103] Zhe Wang, Xiaowen Liu, Jennifer Muther, et al. Top-down mass spectrometry analysis of human serum autoantibody antigen-binding fragments. *Scientific reports*, 9(1):2345, 2019. ISSN 2045-2322. doi: 10.1038/s41598-018-38380-y.
- [104] Rafael D. Melani, Benjamin J. Des Soye, Jared O. Kafader, et al. Next-generation serology by mass spectrometry: Readout of the sars-cov-2 antibody repertoire. *Journal of Proteome Research*, 21(1):274–288, 2022. ISSN 15353907. doi: 10.1021/acs.jproteome.1c00882.
- [105] Jared O. Kafader, Rafael D. Melani, Kenneth R. Durbin, et al. Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nature methods*, 17(4):391–394, 2020. ISSN 1548-7105. doi: 10.1038/s41592-020-0764-5.
- [106] Kazutaka Kitaura, Hiroshi Yamashita, Hitomi Ayabe, et al. Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. *Frontiers in immunology*, 8(May):389, 2017. ISSN 1664-3224. doi: 10.3389/fimmu.2017.00389.
- [107] Albert Bondt, Kelly A. Dingess, Max Hoek, Danique M. H. van Rijswijck, and Albert J. R. Heck. A direct ms-based approach to profile human milk secretory immunoglobulin a (iga1) reveals donor-specific clonal repertoires with high longitudinal stability. *Frontiers in Immunology*, 12:5190, 2021. ISSN 1664-3224. doi: 10.3389/fimmu.2021.789748.
- [108] Noortje de Haan, David Falck, and Manfred Wuhrer. Monitoring of immunoglobulin n- and o-glycosylation in health and disease. *Glycobiology*, 30(4):226–240, 2020. ISSN 1460-2423. doi: 10.1093/glycob/cwz048.
- [109] Qingrong Yan, Maggie Huang, Michael J. Lewis, and Ping Hu. Structure based prediction of asparagine deamidation propensity in monoclonal antibodies. *mAbs*, 10(6):901–912, 2018. ISSN 1942-0862. doi: 10.1080/19420862.2018.1478646.
- [110] Beate Beyer, Manfred Schuster, Alois Jungbauer, and Nico Lingg. Microheterogeneity of recombinant antibodies: Analytics and functional impact. *Biotechnology Journal*, 13(1):1700476, 2018. ISSN 18606768. doi: 10.1002/biot.201700476.

Chapter 1
