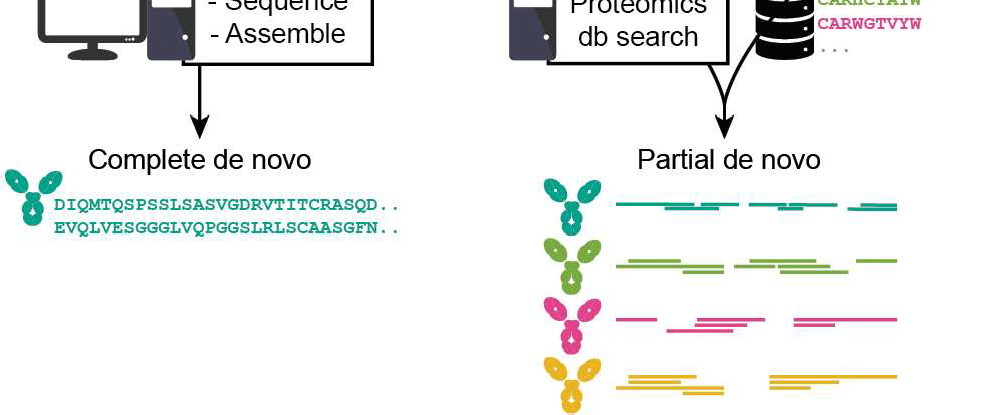
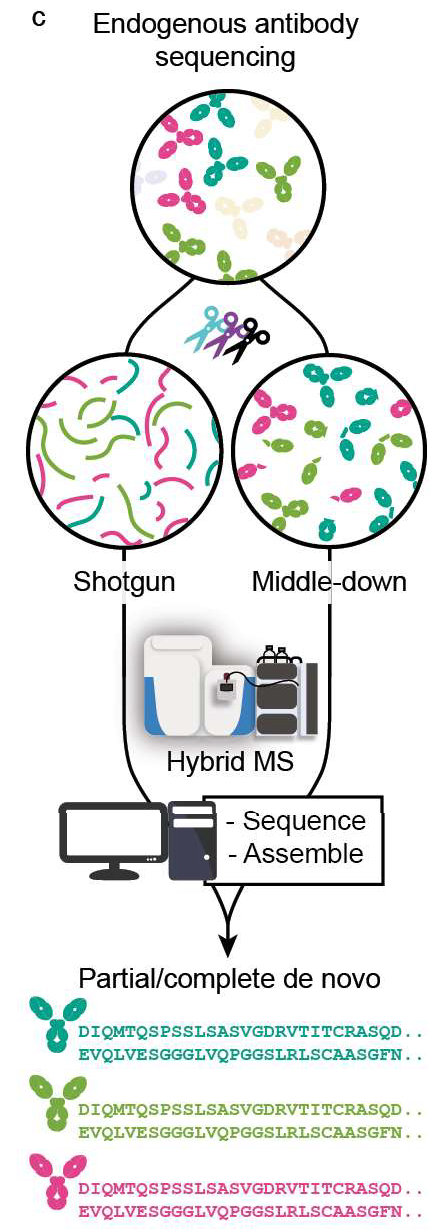
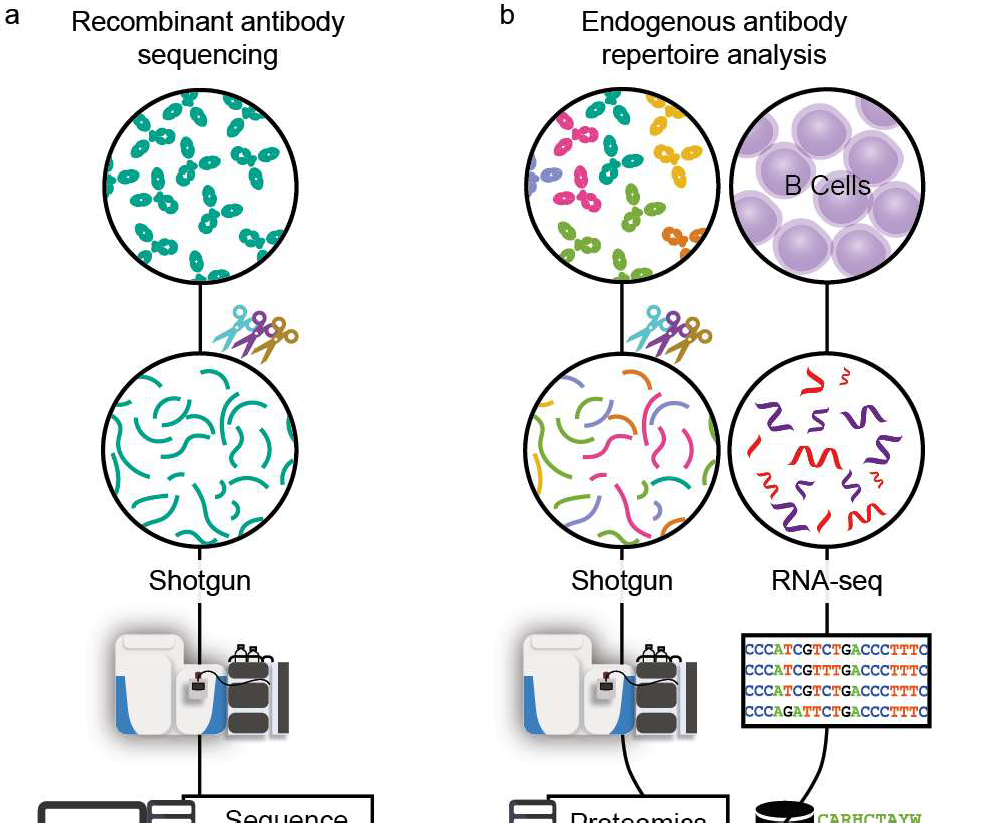
# Introduction



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

**Chapter 1**

**1**

# Prelude - The importance of antibodies

ROUND the time of their initial discovery, antibodies were termed by various il­ lustrious names, such as 'lmmunkorper', 'Amboceptor', and 'Zwischenkorper', among many others. These terms were used more than a century ago to describe substances with antitoxin, lysin, agglutinin, and precipitin activities.(1, 2) Nowadays, the generally accepted term *antibody* refers to secreted immunoglobulins (lgs), whose sequence variety is several orders more diverse than the assortment of their histori­ cal names. Antibodies represent some of the most important molecules in the human immune system. Over the last century, lgs have been intensively studied because of their role in combatting infectious diseases and have taken centre stage for de­ velopment of therapeutics in the last decade.(3-5) Beyond infectious diseases, re­ combinant antibodies are now also developed for cancer, rheumatoid arthritis, and various other pathological conditions.(6) As key entities in the body's defence mech­ anism, circulating antibodies are found in various bodily fluids, such as serum, saliva, milk, the lumen of the gut, and cerebrospinal fluid.(?) 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 lg molecules a human body can generate extends beyond 1015 theoretical sequences,(11, 12) indicating that each antigen may lead to a unique antibody response. These 1015 possible antibody sequences are all unique yet highly alike, posing a serious challenge for their characterization and se­ quencing, 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 re­ ceptor (BCR) sequencing,(13) as is currently the norm, to more directly probe circulat­ ing 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)

A

### Chapter 1

**1**

However, antibodies represent a very special and rather challenging class of pro- teins. 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 effi- ciently into an endogenous humoral antibody discovery and characterization pipeline. In this thesis, I evaluate the role that MS can play in sequencing, identifying, and char- acterizing antibodies, focusing mainly on emerging strategies employed to enable identification and characterization of endogenous neutralizing antibodies.

### Nomenclature, structure, and diversity of antibodies

Humoral human antibodies are complex proteins produced by B cells.(?, 19) Most antibody molecules (e.g. lgGs) 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 con­ nected via disulphide bridges in the hinge region to form the intact antibody. Func­ tionally, 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 Fe or fragment crystallizable)(20) **(Figure 1a).** The Fe is the effector entity of the an­ tibody and can bind to Fe-receptors on immune cells(?) and mediate immune effector responses such as phagocytosis, antibody-dependent cell-mediated cytotoxicity, res­ piratory burst, and cytokine release.(21) In contrast to the fully conserved sequence and structure of the Fe, the Fab is responsible for the vast diversity in recognized anti­ gens and is thus hypervariable. Because there is an endless and constantly evolving pool of pathogens, the antibody repertoire needs to be incredibly diverse and ver­ satile 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:

### Chapter 1

**1**

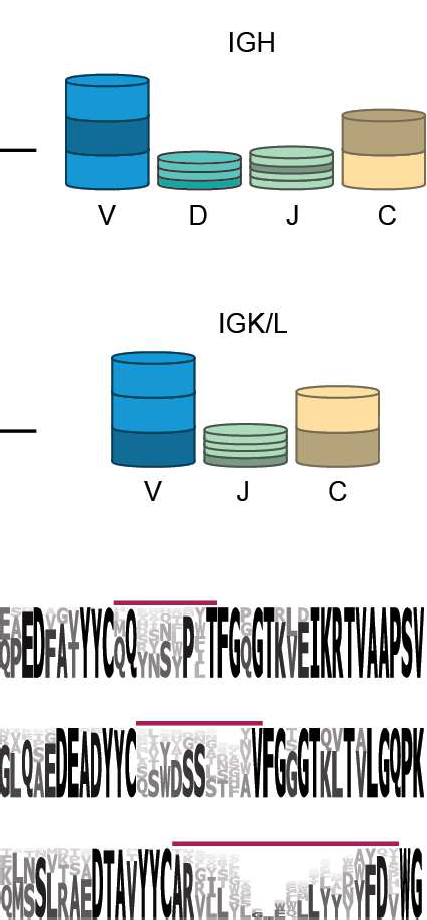
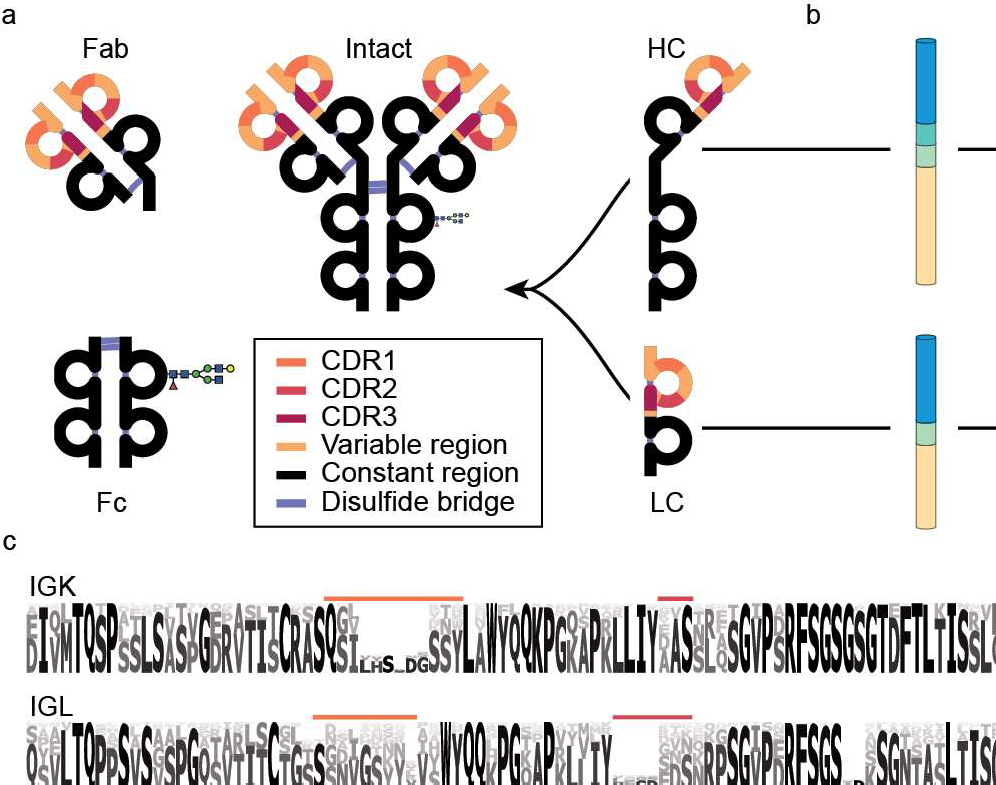
Variable (V), Diversity (D), Joining (J), and Constant (C), with the light chain lacking the D-gene. These genes are encoded in multiple alleles, which can recombine to a staggering number of combinations **(Figure 1**b).(24) The recombination process is also error-prone, leading to insertions and deletions 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 hypervariabil­ ity, 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), sep­ arated 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 se­ quences is of the utmost importance for antibody discovery. The Fe part of lgs is used to classify antibodies into one of 5 classes: lgA, lgD, lgE, lgG, and lgM. Some of these classes are divided further into subclasses denoted by numbers, e.g., lgG1- 4 or lgA1 and lgA2. 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 lgG1, the most abundant antibody subclass in serum, and the predominantly used subclass for biotherapeutic development. Still, concerning *de nova* sequencing by MS, different lg classes and subclasses pose similar challenges and opportunities.

### 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

1



**Figure 1: Nomenclature, structure, and diversity of lgG1 antibodies.** (a) Nomenclature and protein fragments of an lgG1 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 Fe. Other lgG subclasses vary in their heavy chain constant region (Fe) 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 silica generated sequences of lg kappa (IGK) and lambda (IGL) light chains and lg heavy chain (IGH) from the international lmMunoGeneTics (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 frag­ mentation adhere to highly specific rules, peptides and their gas-phase fragment ions can be predicted. Consequently, peptides and their parent proteins are iden-

### Chapter 1

**1**

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 nova* 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 sam­ ple. 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 complex­ ity of antibody repertoire or following changes in abundance of specific clones.(30) When applied to *de nova* sequencing, the precursor mass knowledge can help de­ termining the light and heavy chain pairing or sequence prediction accuracy in BU sequencing.(31) In addition, both denatured and native antibodies can also be frag­ mented 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)

**Chapter 1**

**1**

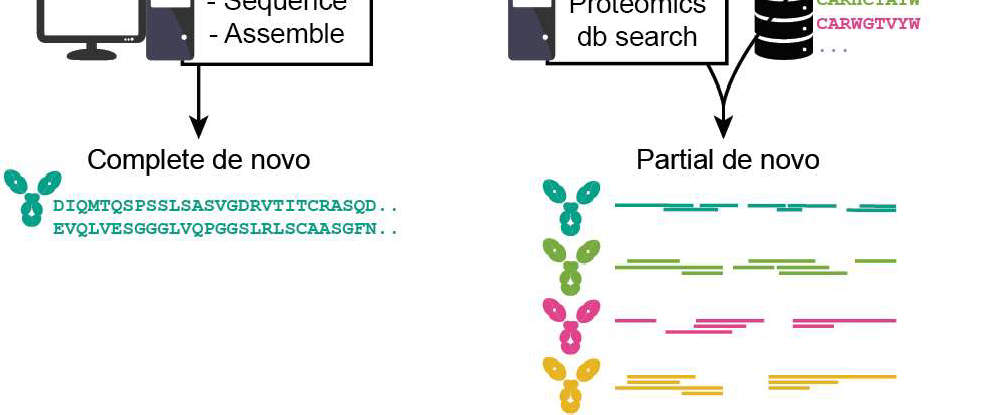
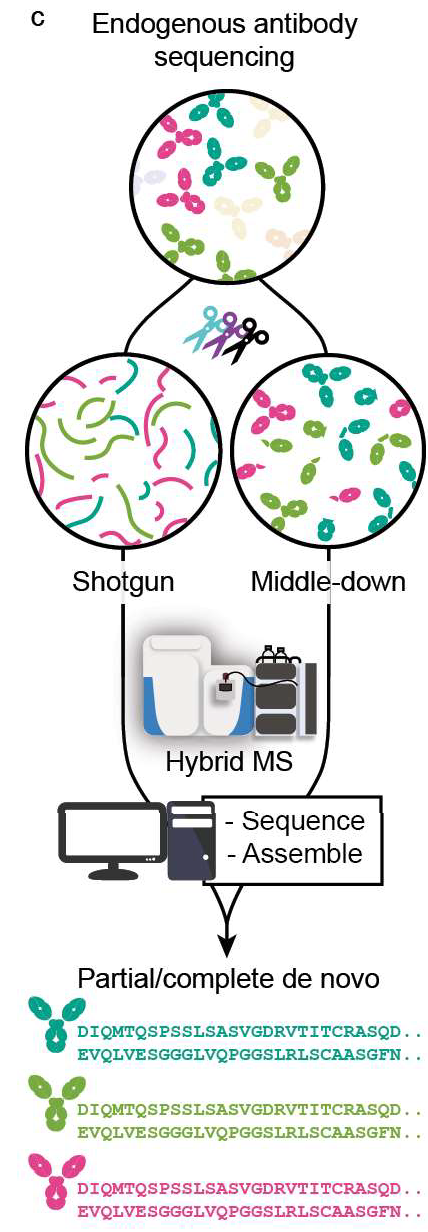
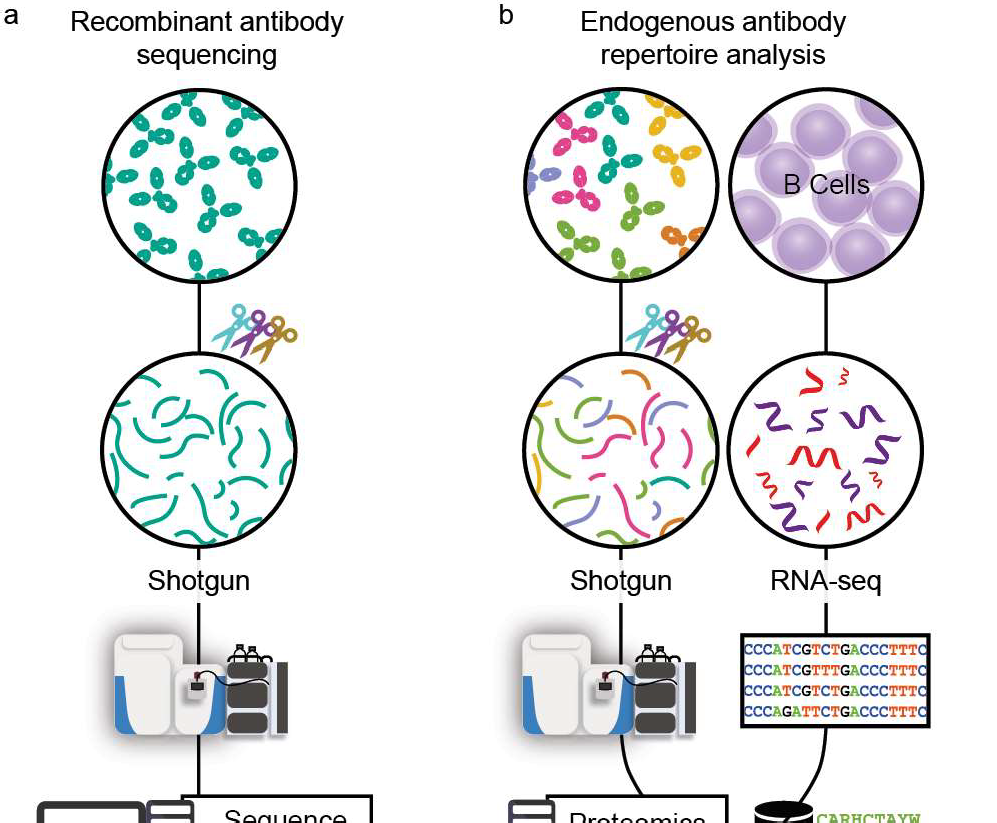
# The emerging role of mass spectrometry in anti- body discovery

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 there­ after 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 ex­ pensive 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 suc­ cessfully 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 in­ fectious 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 prepa­ ration, MS and liquid chromatography instrumentation, and data analysis.(14, 15)

### Chapter 1

**1**



**Figure 2: Three approaches in MS-based antibody sequencing.** (a) Recombinant antibody sequenc­ ing 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, provid­ ing multiple overlapping short sequence reads. After liquid chromatography-mass spectrometry (LC-MS) measurement, the spectra can be processed by several different de nova 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 gen­ eration of the personalized database, a high throughput of the LC-MS is possible.(43) While not strictly de nova since only hits from the sequence database are identified, it is a powerful tool for antibody reper­ toire analysis. (c) Endogenous antibody sequencing cannot rely on BU MS alone, as direct sequencing of endogenous humeral 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 anti­ body 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 deter­ mine antibody sequences. These three pillars are primarily classified by the source of sample material and the attainable sequencing information. The first strategy ap­**

**plies to highly purified recombinant antibodies that are now amenable for full** se-

### Chapter 1

**1**

quencing with BU MS, often by combining several different proteases and advanced algorithms. Second, hybrid approaches have been introduced for analysing endoge- nous antibody repertoires by combining MS-based techniques with genomics or tran­ scriptomics, e.g., whole genome sequencing or BCR sequencing, ideally from the same donor. The third set of techniques encompasses several MS-based *de nova* 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.

### 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.

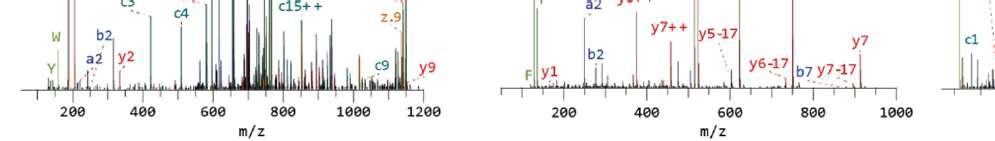
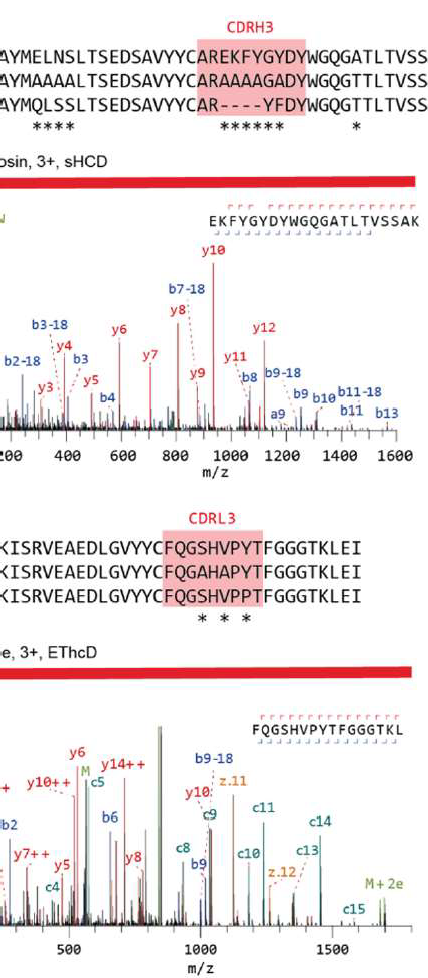
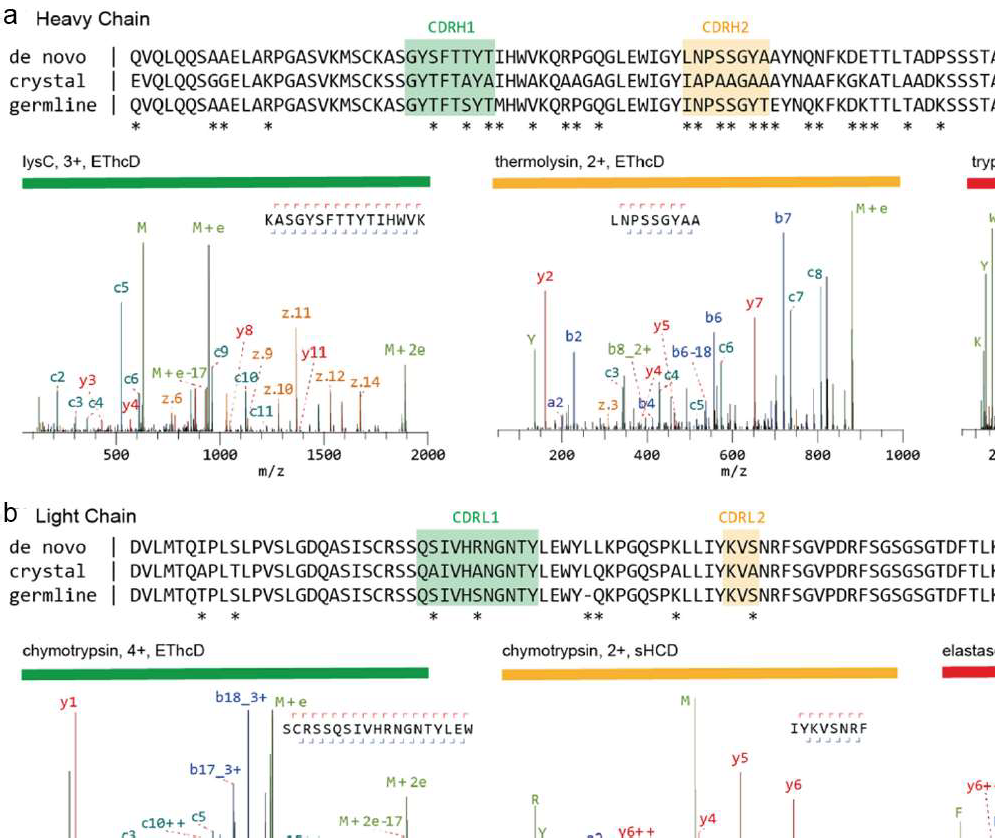
**Shotgun, bottom-up strategies used for sequencing of highly purified mAbs**

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 rep­ resent the most popular type of proteomics experiments. In contrast to most shotgun proteomics experiments, *de nova* 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 pro­ tease 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

### Chapter 1

**1**

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, emphde novo sequencing of antibodies thus requires other ap­ proaches. Several methods to generate complete and deep sequence coverage by overlapping peptides have been introduced. For example, a shortened protease incu­ bation time was successfully used to increase the number of peptides carrying missed cleavage sites.(44) Some proteases generate a high number of overlapping pep­ tides 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 con­ trolled 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 nova* 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 nova* se­ quencing solutions, such as ALPS/PeaksAB,(27) GenoMS,(46) SuperNovo(16) and Champs(48) are quite successful in obtaining full sequence coverage of highly pu­ rified antibodies. To determine the antibody sequences *de nova,* 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 pro­ teases. While complicating sample preparation and increasing the required amount, such multi-protease approaches are advantageous for *de nova* sequencing by alle­ viating the sequence assembly problem.



**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 nova sequence derived by MS is shown on top, alongside the previ­ ously published sequence used 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 align­ ments, 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 intacUcharge 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 gen­ erate arrays of adjacent fragments that reveal the amino acid sequence, often re­ ferred 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

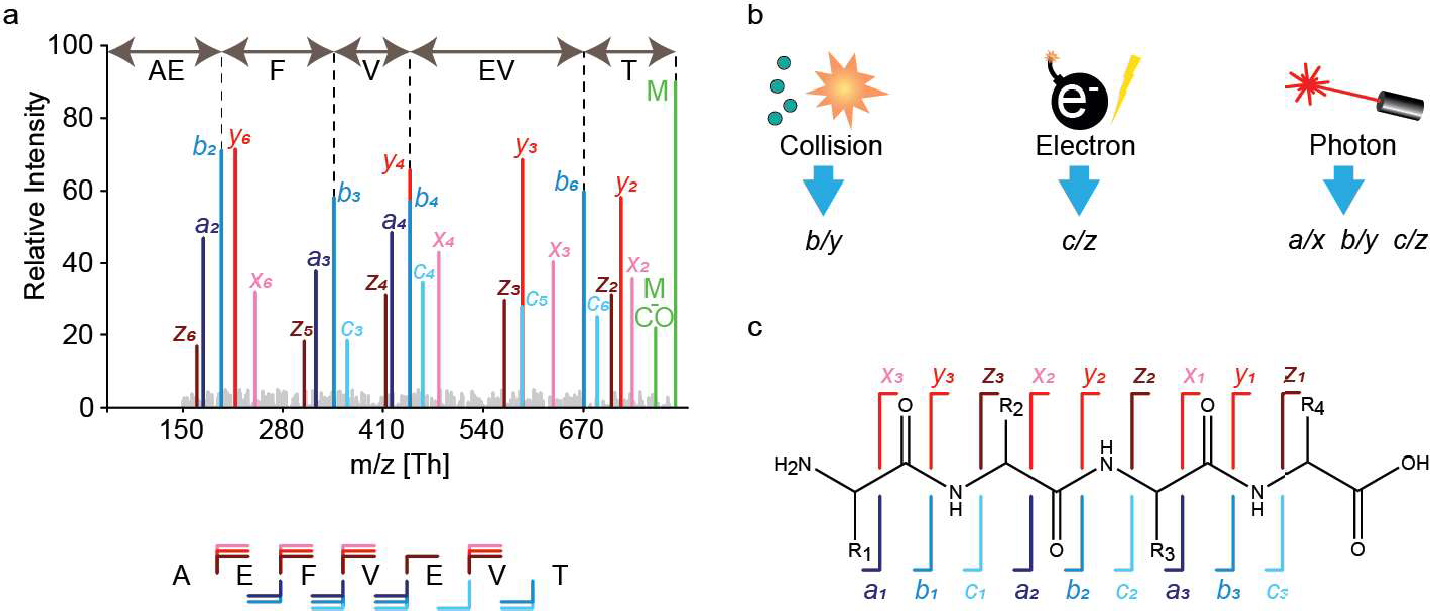
#### Chapter 1

**1**

resolution.(49) Since there is no universal fragmentation method that can produce un­ interrupted fragment ion ladders for all possible peptides, it is highly advantageous to use various fragmentation methods with distinct mechanisms and specificities to com­ plement 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 complemen­ tary. These specificities stem from the unique ion activation mechanisms employed by each method. In collision-based techniques, energy is deposited to the multi­ ply 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 phosphory­ lation and sialyation. In electron-based techniques (ECO/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 con­ trast 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-Ca 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, *bly, clz),* 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 nova* 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-



**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 *bly* ions. Electron based dissociation (ECO/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 be­ long. 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 nova* protein sequencing by combining CID and ECO 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 frag­ mentation 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

### Chapter 1

**1**

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 strate­ gies have made *de nova* 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 fragmenta­ tion techniques to obtain unambiguous sequence reads, which entails longer sample preparation time, the requirement of larger sample amounts, and extensive data ac­ quisition.

**Homology-aided de novo sequencing of antibodies**

To identify peptides and proteins, shotgun proteomics experiments rely on match­ ing 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 pro­ teins like antibodies. Instead, homologous sequences, primarily derived from ge­ nomic or transcriptomic experiments, can be used. For antibodies, the genes encod­ ing 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 homolo­ gous sequences can facilitate verification or guide predictions of *de nova* 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 ap­ plicable because they require exact mass matching of fragments, and a single amino acid mutation can prevent identification. Instead, error-tolerant fragment matching al­ gorithms, either based on sequence alignments or subsequence (i.e., sequence tag) extractions, can use homologous databases to score experimentally determined se­ quences. An example of a homology-aided approach is searching BU MS data from

#### Chapter 1

**1**

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 nova* 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 nova* sequencing algorithms are also advantageous in identifying erroneous *de nova* peptide reads by comparing them against homologous sequences. In addition, they can be used as a germline tem- plate to aid in the assembly of *de nova* 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 nova* 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 chal­ lenges. First, compared to peptides, intact proteins sometimes ionize less efficiently, and liquid-chromatography-based separation of peptides is more established and ef­ ficient than separation of intact proteins. Furthermore, in MS analysis, mass accu­ racy and resolution typically diminish with increasing molecular weight, even when using the latest high-resolution mass spectrometers.(29, 70, 71) In addition, full se­ quence 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 nova* sequencing of antibodies. However, more recently, several advances in the field resulted in relatively high sequence coverages, reported for re­ combinant mAbs with available reference sequences.(35, 72-74) Protein-centric ap­ proaches, termed TD MS,(32) can provide additional valuable information, including

#### Chapter 1

**1**

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 spe­ cific proteases, such as lgdE (commercially termed FabALACTICA), which cleaves above the hinge region of lgG1, 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 lgG subclasses.(34) Such strate­ gies 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 anal­ ysed. 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 cov­ erage 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 ECO 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 ob­ tained. The resulting fragmentation spectrum contained not only the multiply charged backbone fragmentation products but also the intact light chain, ejected from the an­ tibody by fragmentation of the intermolecular disulphide bridge, providing information on the light and heavy chain pairing **(Figure 5c).** These and many other studies cul­ minated in a large joint effort by the Consortium for Top-Down Proteomics, wherein they comprehensively described available approaches, techniques, and instrumen­ tation for the analysis of recombinant mAbs.(18) Electron-based fragmentation of

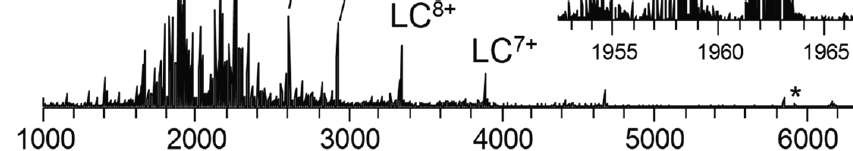
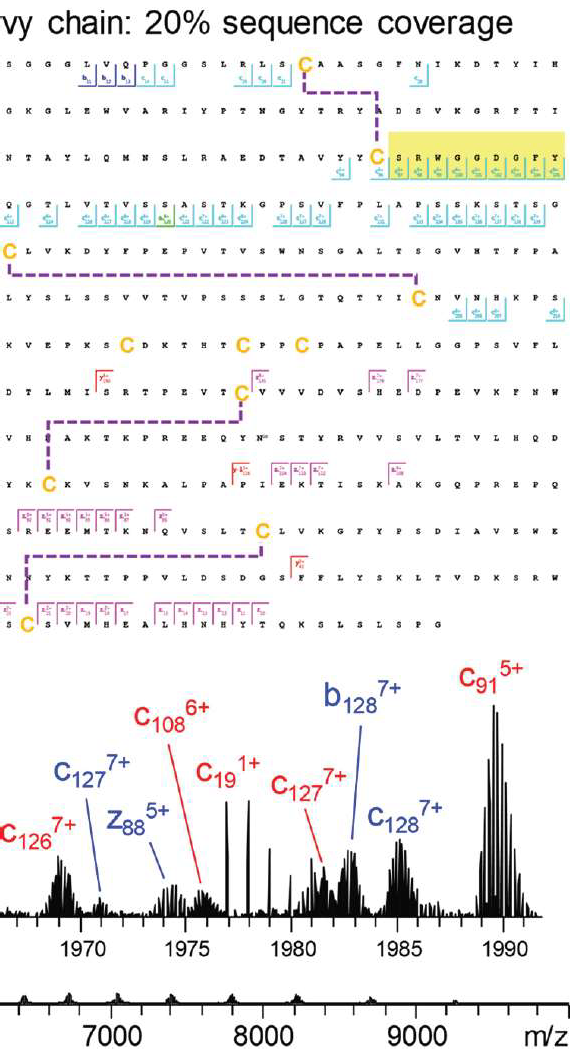
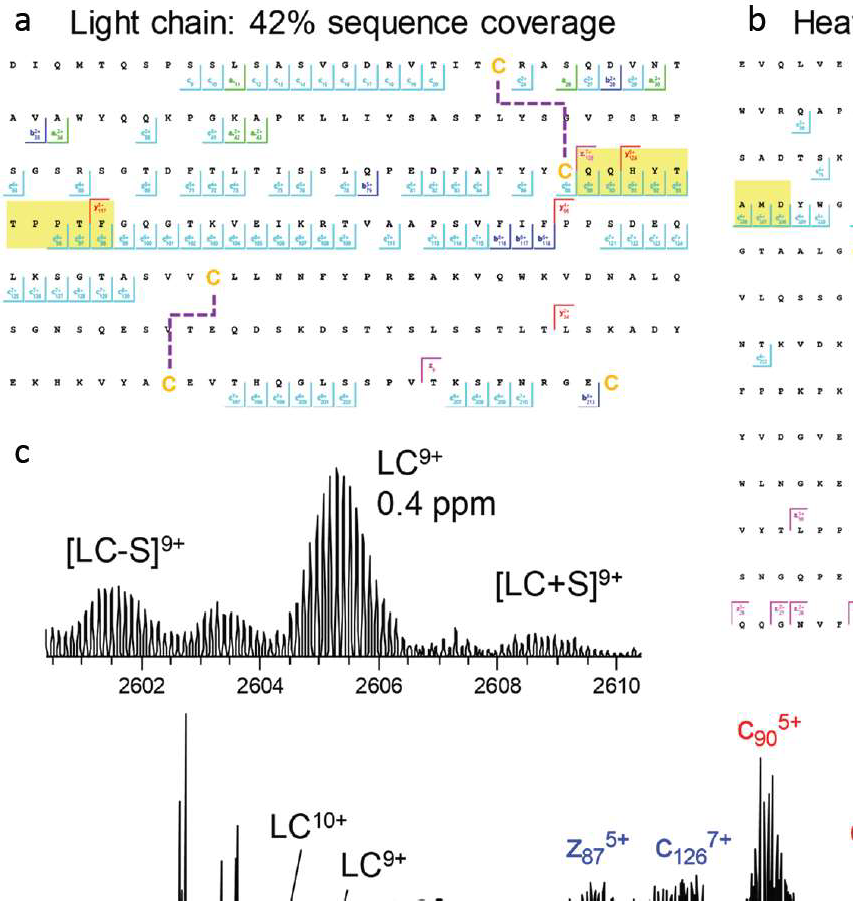
#### Chapter 1

**1**

intact protein ions holds great potential for mAb sequencing. Several recent stud- ies showed that these methods consistently yielded nearly uninterrupted c-ion lad- ders spanning the CDR3, which is paramount to antigen binding.(30, 75, 76, 83, 84) These studies also demonstrated for various antibody isotypes (lgG1-4 and lgA1) that electron-based fragmentation methods consistently provide fragments contain- ing 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 ldeS enzyme), and Fab molecules (produced with lgdE or Operator enzymes), showing that reduc- ing antibody complexity through the removal of the Fe 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 anti- body 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 po- tential benefits that protein-centric approaches bring to the *de nova* analytical toolbox are hard to neglect. While it is still nearly impossible to fully *de nova* sequence intact mAbs, protein-centric sequencing can be combined with peptide-centric methods in a hybrid MS approach, providing complementary information substantially advanc- ing 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).**

### Chapter 1

1



**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 spec­ trum (c) for the 25+ charge state of intact Trastuzumab with insets displaying the zoomed in region con­ taining 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 pre­ sented above result in extended datasets that are not feasible for manual interpreta­ tion. Thus, development of dedicated software tools for data interpretation is essen­ tial. With regards to BU MS data, presently, two popular software suites are tailored towards *de nova* 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 se­ quence database like IMGT to make a complete *de nova* sequence prediction based on the BU MS data. More specifically, the software iteratively screens predicted pep-

#### Chapter 1

**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 anti- bodies. Novel software solutions for *de nova* antibody sequencing are emerging and advancing in parallel with improvements in experimental design and instrumentation. The fast development of new *de nova* sequencing strategies encourages the devel- opment of new software solutions and improvement of already established tools and requires adaptable software to accommodate the frequent and considerable shifts in *de nova* 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 sequenc­ ing**

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 nova* predicted peptide reads to fill the complete sequence. Their analysis makes use of external BU *de nova* sequencing software, PEAKS,(86) and was tested on protein mixtures. TBNovo has not achieved widespread adoption, perhaps due to the soft­ ware'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

### Chapter 1

**1**

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 recom­ binant mAb samples are typically available in milligram quantities, polyclonal antibody samples are often derived from clinical samples and thus will only be available in lim­ ited quantities. Because the median concentration of individual clones in plasma is

~1 µg/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 lg sequences. Additionally, in complex endogenous poly­ clonal 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.

### 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 strate­ gies. Using a multi-omics approach, for instance, by supplementing BU MS data with genomics or transcriptomics data derived from the same donor, allows bypass­ ing some challenging aspects of genuine *de nova* sequencing, albeit at the cost of a more complex, labour- and data-intensive workflow **(Figure 2c).** Presently, direct *de nova* sequencing of antibodies from a complex mixture is still a tremendous chal-

#### Chapter 1

**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.

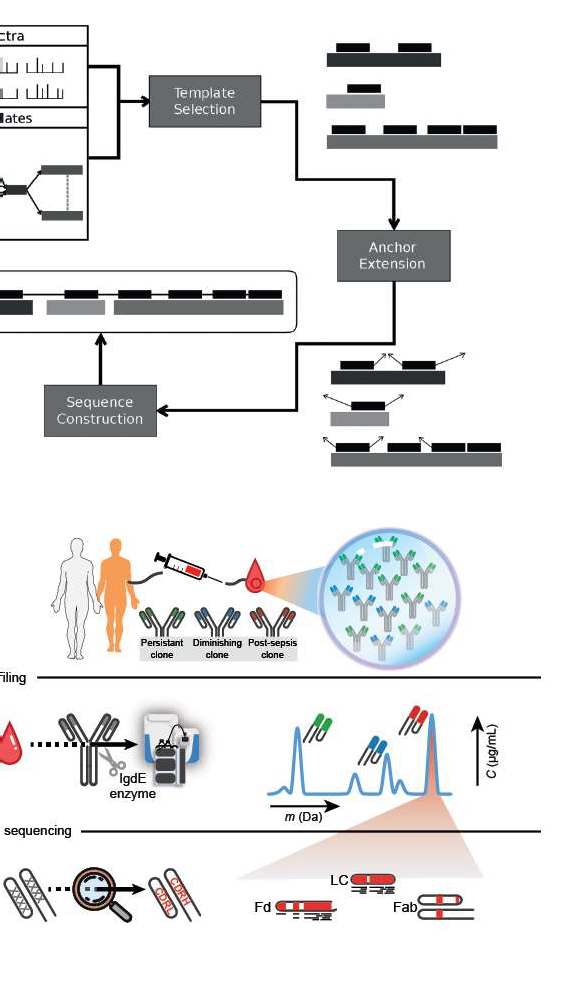
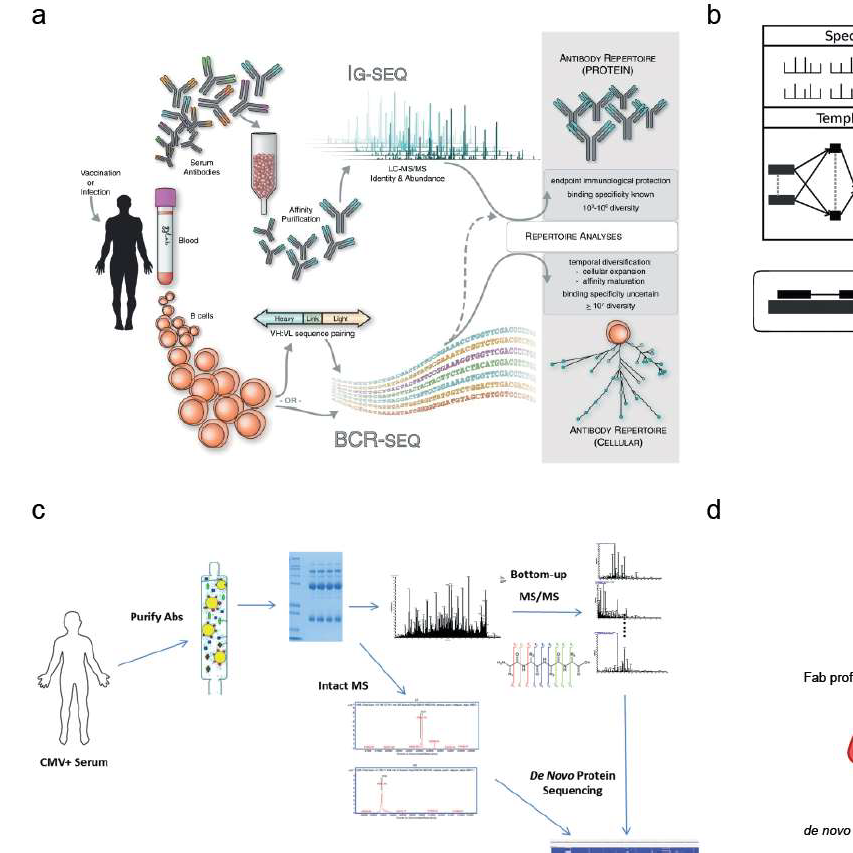
**lg-seq**

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 lg-seq method pioneered by Lavinder et al.(89) in the Georgiou lab ap­ plies 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 nova* sequencing **(Figure Ga).** 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 lgG 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 lon­ gitudinal monitoring of influenza antibodies over multiple years. Monitoring the ef­ fects 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 lg­ 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, lg-seq presents one of the most effi­ cient and successful approaches to analyse and identify clones in lg repertoires and

### Chapter 1

**1**

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



**Figure 6: Selected recent approaches aiming towards MS-based de novo sequencing of serum antibodies.** (a) In lg-seq(89) a personalized database generated by BCR sequences is used to iden­ tify 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 poly­ clonal 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 quanti­ fies intact masses of Fabs to provide a view of the lgG1 clonal repertoire, enabling to quantify and monitor individual clones. Abundant serum clones are identified by using BU and MD MS data iteratively to gen­ erate full lgG *de nova* sequences. Figure adapted from Bondi et al.(30)

**Alternative proteogenomics approaches**

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

### Chapter 1

**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 flex- ibility through user-defined constraints. In addition, users can define how the tem- plate 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 hy- brid 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 anti- bodies 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)

### Protein-centric sequencing of endogenous antibodies

Some attempts have emerged aiming at novel antibody discovery by MS-based se­ quencing alone, directly from serum samples or other liquid biopsies, circumventing the need for genomics/transcriptomics data (multi-omics approaches). Above we re­ viewed 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 nova* sequences of endogenous antibody clones by combining different mass spectrometric techniques, as discussed further below.

### Chapter 1

**1**

**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 affin­ ity 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 anti­ gen co-occur. An example of a capturing method whereby additional intact mass data was used to derive *de nova* sequences was described by Guthals et al.(31) Follow­ ing affinity purification of antibodies from the serum of a cytomegalovirus-exposed individual, using the glycoprotein B antigen, both intact mass and BU MS measure­ ments 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 se­ quence derived using the BU MS data **(Figure Ge).** 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, di­ verging 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 nova* 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.

#### Chapter 1

**1**

**Antibody profiling and sequencing in polyclonal mixtures**

While it is still not possible to *de nova* sequence entire serum antibody repertoires, recent advances in LC-MS of intact proteins enabled detecting and resolving sin­ gle clones from complex antibody mixtures. For instance, developments have been made that specifically profile intact light chains from serum, even providing partial se­ quence information by using MD MS. Impressively these studies successfully demon­ strate 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 lg overexpression in serum (mono-gammopathy) such as multiple myeloma. Nonetheless, these studies demonstrated that detection and characterization of in­ dividual 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 nova* 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 lg­ 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 sam­ ple 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 met­ rics are correlated to the ELISA-based antibody titer and neutralization efficiency, to verify their accuracy. In a recent study, Bandt et al.(30) used an approach to generate

### Chapter 1

**1**

Fab fragments exclusively from the entire lgG1 repertoire. They were able to longitu­ dinally profile lgG1 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 lgG1 Fab clones per donor and showed that most clones persist over multiple months of sampling. Contrary to widely held belief, they showed that the lgG1 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 nova* sequence a single highly abundant clone in one of the donors without the aid of antigen-specific capture. The *de nova* sequencing was achieved by a combina­ tion of protein-centric sequencing using ETD, and a BU MS approach using multiple proteases for digestion. First, closely matching light and heavy chain germline tem­ plates 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 nova* sequencing of clones directly from serum is feasible, although still arduous and limited to specific cases **(Figure 6d).** Notably, the deter­ mined sequences contained more mutations (compared to germline sequences) than expected from the reported rates from BCR sequencing studies,(106) which is indica­ tive of potential discrepancies between protein-level and gene-level sequencing. This first attempt focused exclusively on lgG1, by using an lgG1-specific protease to gen­ erate the Fab fragments. In another work, Bondt et al.(107) extended their method to lgA1 by using a protease specific to the O-glycans present in lgA1 hinge region to generate Fab fragments, albeit now exclusively from lgA1. Overall, they showed that

- similar to serum lgG1 - just a handful of clones dominates the secretory lgA1 profile of human milk. Using a somewhat comparable approach, Dupre et al.(102) analysed isolated light chains from the urine of a patient affected by multiple myeloma. They assembled *de nova* 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

### Chapter 1

**1**

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

### 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 lgG1 are thoroughly described, other subclasses, notably lgG2, appear to occur as structural isomers induced by different disulphide-bridge patterns. These PTMs and disulphide bridges become even more pronounced in lgA and lgM, 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.

**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.

**Chapter 1**

# References

1

1. ES London. Der gegenwartige stand der lehre von den cy­ tolysinen und die cytolytische theorie der immunitat. *Zentralbl.*

*f. Bakteriol., Parasitenk. u. lnfektionskrankh., Abt. II, Bd,* 32, 1902.

1. Jean Lindenmann. Origin of the terms 'antibody' and 'antigen'.

*Scandinavian Journal of Immunology,* 19(4):281-285, 1984. ISSN 0300-9475. doi: 10.1111lj.1365-3083.1984.tb00931.x.

1. 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: 10.1074ljbc.rev120.010181.
2. Matthew I J Raybould, Claire Marks, Alan P Lewis, et al.

Thera-sabdab: the therapeutic structural antibody database.

*Nucleic acids research,* 48(1):383-388, 2020. ISSN 1362-

4962. doi: 10.1093lnarlgkz827.

1. Helene Kaplon and Janice M. Reichert. Antibodies to watch in 2021. *mAbs,* 13(1), 2021. ISSN 19420870. doi: 10.10801

19420862 .2020.18604 76.

1. Surjit Singh, Nitish K. Kumar, Pradeep Dwiwedi, et al. Mon­ oclonal antibodies: A review. *Current Clinical Pharmacol­ ogy,* 13(2):85-99, 2018. ISSN 15748847. doi: 10.21741

1574884712666170809124728.

1. Harry W. Schroeder and Lisa Cavacini. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunol­ ogy,* 125(2):41-52, 2010. ISSN 00916749. doi: 10.1016lj.jaci.

2009.09.046.

1. 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: 10.1126lscience.aad5788.
2. 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: 10.1126lscience.

aad5224.

[1OJ Rut Valgardsdottir, Irene Cattaneo, Gavino Napolitano, et al. Identification of human sars-cov-2 monoclonal antibodies from convalescent patients using ebv immortalization. *Antibod­* ies, 10(3):26, 2021. ISSN 2073-4468. doi: 10.33901

antib10030026.

1. H W Schroeder Jr. Similarity and divergence in the devel­ opment and expression of the mouse and human antibody repertoires. *Dev Comp lmmunol,* 30(12):119-135, 2006. doi: 10.1016lj.dci.2005.06.006.
2. Bryan Briney, Anne lnderbitzin, Collin Joyce, and Den­ nis R. Burton. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature,* 566 (7744):393-397, 2019. ISSN 1476-4687. doi: 10.10381 s41586-019-0879-y.
3. 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.

1. 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: 10.1038lnrg3356.

1. Ruedi Aebersold and Matthias Mann. Mass-spectrometric ex­ ploration of proteome structure and function. *Nature,* 537 (7620):347-355, 2016. ISSN 0028-0836. doi: 10.10381

nature19949.

1. K. liker Sen, Wilfred H Tang, Shruti Nayak, et al. Automated antibody de novo sequencing and its utility in biopharmaceu­ tical discovery. *Journal of the American Society for* Mass *Spectrometry,* 28(5):803-810, 2017. ISSN 1879-1123. doi: 10.1007ls13361-016-1580-0.
2. Weiwei Peng, Matti F. Pronker, and Joost Snijder. Mass spectrometry-based de novo sequencing of monoclonal an­ tibodies using multiple proteases and a dual fragmenta­ tion scheme. *Journal of Proteome Research,* 20(7):3559- 3566, 2021. ISSN 1535-3893. doi: 10.1021lacs.jproteome. 1c00169.

#### Chapter 1

1. Kristina Srzentic, Luca Fornelli, Yury 0. 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.

1

1. 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.
2. R R Porter. The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. *Biochemical Journal,* 73(1):119-127, 1959. ISSN 0306-3283. doi: 10.1042/bj0730119.
3. Andrew B. Herr, Edward R. Ballister, and Pamela J. Bjorkman.

Insights into iga-mediated immune responses from the crys­ tal structures of human fcalphari and its complex with iga1- fc. *Nature,* 423(6940):614-20, 2003. ISSN 0028-0836. doi:

10.1038/nature01685.

1. Jr Charles A Janeway, Paul Travers, Mark Walport, and Mark J Shlomchik. The generation of diversity in immunoglob­ ulins. In *lmmunobiology: The Immune System in Health and* Disease. New York (NY), 5 edition, 2001. ISBN 0-8153-3642- X.
2. 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.

1. DJ Jeske, J Jarvis, C Milstein, and J D Capra. Junctional di­ versity is essential to antibody activity. *Journal of immunology,* 133(3):1090-1092, 1984. ISSN 0022-1767.
2. Marie Paule Lefranc. lmgt, the international immunogenetics database®. *Nucleic Acids Research,* 31(1):307-310, 2003. ISSN 03051048. doi: 10.1093/nar/gkg085.
3. Ruedi Aebersold and Matthias Mann. Mass spectrometry­ based proteomics. *Nature,* 422(6928):198-207, 2003. ISSN

0028-0836. doi: 10.1038/nature01511.

1. Ngoc Hieu Tran, M Ziaur Rahman, Lin He, et al. Complete de novo assembly of monoclonal antibody sequences. *Sci­ entific Reports,* 6(1):31730, 2016. ISSN 2045-2322. doi: 10.1038/srep31730.
2. Adrian Guthals, Karl R Glauser, and Nuno Bandeira. Shot­ gun protein sequencing with meta-contig assembly. *Molecu­ lar and Cellular Proteomics,* 11(10):1084-1096, 2012. ISSN 15359476. doi: 10.1074/mcp.m111.015768.
3. Daniel P. Donnelly, Catherine M. Rawlins, Caroline J. De­ Hart, et al. Best practices and benchmarks for intact pro­ tein analysis for top-down mass spectrometry. *Nature Meth­ ods 2019 16:7,* 16(7):587-594, 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0457-0.
4. Albert Bandt, Max Hoek, Sem Tamara, et al. Human plasma igg1 repertoires are simple, unique, and dynamic. Ce// Sys­ *tems,* 12(12):1131-11435, 2021. ISSN 24054720. doi: 10. 1016/j.cels.2021.08.008.
5. Adrian Guthals, Yutian Gan, Laura Murray, et al. De novo ms/ms sequencing of native human antibodies. *Journa/ofPro­ teome Research,* 16(1):45-54, 2017. ISSN 1535-3893. doi: 10.1021/acs.jproteome.6b00608.
6. Timothy K. Toby, Luca Fornelli, and Neil L. Kelleher.

Progress in top-down proteomics and the analysis of pro­ teoforms. *Annual Review of Analytical Chemistry,* 9 (1):499-519, 2016. ISSN 1936-1327. doi: 10.1146/

annurev-anchem-071015-041550.

1. 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.
2. Bjorn P. Johansson, Oonagh Shannon, and Lars Bjtirck.

Ides: A bacterial proteolytic enzyme with therapeutic poten­ tial. *PLoS ONE,* 3(2):1692, 2008. ISSN 1932-6203. doi:

10.1371/joumal.pone.0001692.

1. Luca Fomelli, Daniel Ayoub, Konstantin Aizikov, Alain Beck, and Yury O Tsybin. Middle-down analysis of monoclonal anti­ bodies with electron transfer dissociation orbitrap fourier trans­ form mass spectrometry. *Analytical Chemistry,* 86(6):3005- 3012, 2014. ISSN 0003-2700. doi: 10.1021/ac4036857.
2. 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 necro­ sis factor. *Molecular and Cellular Biology,* 9(3):1165-1172, 1989. ISSN 0270-7306. doi: 10.1128/mcb.9.3.1165-1172.

1989.

### Chapter 1

1. Paul Carter, L. Presta, Cornelia M Gorman, et al. Human­ ization of an anti-p185her2 antibody for human cancer ther­ apy. *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.

1

1. Naoya Tsurushita, Paul R. Hinton, and Shankar Kumar. De­ sign of humanized antibodies: From anti-tac to zenapax. *Methods,* 36(1):69-83, 2005. ISSN 10462023. doi: 10.1016/ j.ymeth.2005.01.007.
2. Leila Khoja, Marcus O Butler, S Peter Kang, Scott Ebbing­ haus, and Anthony M Joshua. Pembrolizumab. *Journal for lmmunoTherapy of Cancer,* 3(1):36, 2015. ISSN 2051-1426. doi: 10.1186/s40425-015-0078-9.
3. 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.

1. LG Presta, SJ Lahr, R L Shields, et al. Humanization of an antibody directed against ige. *Journal of immunology,* 151(5): 2623-2632, 1993. ISSN 0022-1767.
2. 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.
3. 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/nbl.2782.

1. Denis Morsa, Dominique Baiwir, Raphael La Rocca, et al.

Multi-enzymatic limited digestion: The next-generation se­ quencing for proteomics? *Journal of Proteome Research,* 18 (6):2501-2513, 2019. ISSN 1535-3893. doi: 10.1021/acs.

jproteome.9b00044.

1. Nuno Bandeira, Victoria Pham, Pavel Pevzner, David Arnott, and Jennie R Lill. Beyond edman degradation: Automated de nova protein sequencing of monoclonal antibodies. *Nature Biotechnology,* 26(12):1336-1338, 2008. ISSN 1087-0156. doi: 10.1038/nbt1208-1336.
2. Natalie E Castellana, Victoria Pham, David Amott, Jennie R Lill, and Vineet Bafna. Template proteogenomics: Sequencing whole proteins using an imperfect database. *Molecular and CellularProteomics,* 9(6):1260-1270, 2010. ISSN 15359476.

doi: 10.1074/mcp.m900504-mcp200.

1. Alon Savidor, Rotem Barzilay, Dalia Elinger, et al. Database­ independent protein sequencing (dips) enables full-length de novo protein and antibody sequence determination. *Molecular andcellularproteomics: MCP,* 16(6):1151-1161, 2017. ISSN

1535-9484. doi: 10.1074/mcp.0116.065417.

1. Xiaowen Liu, Yonghua Han, Denis Yuen, and Bin Ma. Auto­ mated protein (re)sequencing with ms/ms and a homologous database yields almost full coverage and accuracy. *Bioin­ formatics,* 25(17):2174-2180, 2009. ISSN 1460-2059. doi:

10.1093lbioinformatics/btp366.

1. Lidong He, Chad R. Weisbrod, and Alan G. Marshall. Protein de nova sequencing by top-down and middle-down ms/ms: Limitations imposed by mass measurement accuracy and gaps in sequence coverage. *lntemational Journal of Mass*

*Spectrometry,* 427:107-113, 2018. ISSN 13873806. doi:

10.1016/j.ijms.2017.11.012.

1. Luis A. Macias, Ines C. Santos, and Jennifer S. Brodbelt.

Ion activation methods for peptides and proteins. *Analyti­ cal Chemistry,* 92(1):227-251, 2020. ISSN 0003-2700. doi:

10.1021/acs.analchem.9b04859.

1. 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.

1. 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.

1. John E. P. Syka, Joshua J. Coon, Melanie J. Schroeder, Jef­ frey Shabanowitz, and Donald F. Hunt. Peptide and pro­ tein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sci­ ences,* 101(26):9528-9533, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0402700101.

#### Chapter 1

1. Roman A. Zubarev, David **M.** Horn, Einar K. Fridriksson, et al.

**1**

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.

1. Scott A. Mcluckey and James L. Stephenson. Ion/ion chem­ istry 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.
2. Jennifer S. Brodbelt, Lindsay J. Morrison, and Ines Santos. Ul­ traviolet photodissociation mass spectrometry for analysis of biological molecules. *Chemical Reviews,* 120(7):3328-3380, 2020. ISSN 0009-2665. doi: 10.1021/acs.chemrev.9b00440.
3. Yongsheng Xiao, Malgorzata M. Vecchi, and Dingyi Wen.

Distinguishing between leucine and isoleucine by integrated le-ms analysis using an orbitrap fusion mass spectrome­ ter. *Analytical Chemistry,* 88(21):10757-10766, 2016. ISSN

15206882. doi: 10.1021/acs.analchem.6b03409.

1. Frank Kjeldsen, Kim F. Haselmann, Esben S. S0rensen, and Roman A. Zubarev. Distinguishing of ile/leu amino acid residues in the pp3 protein by (hot) electron capture dissoci­ ation in fourier transform ion cyclotron resonance mass spec­ trometry. *Analytical Chemistry,* 75(6):1267-1274, 2003. ISSN

0003-2700. doi: 10.1021/ac020422m.

1. D M Horn, R A Zubarev, and F W Mclafferty. Automated de nova sequencing of proteins by tandem high-resolution mass spectrometry. *Proceedings of the National Academy of Sci­ ences,* 97(19):10313-10317, 2000. ISSN 0027-8424. doi: 10.1073/pnas.97.19.10313.
2. Adrian Guthals, Kar1 R. Glauser, Ari M. Frank, and Nuno Ban­ deira. Sequencing-grade de nova analysis of ms/ms triplets (cid/hcd/etd) from over1apping peptides. *Journal of Proteome Research,* 12(6):2846-2857, 2013. ISSN 15353893. doi: 10.1021/pr400173d.
3. Kira Vyatkina. De nova sequencing of top-down tandem mass spectra: A next step towards retrieving a complete protein se­ quence. *Proteomes,* 5(1):6, 2017. ISSN 2227-7382. doi: 10.3390/proteomes5010006.
4. Ritendra Datta and Marshall Bern. Spectrum fusion: Using multiple mass spectra for de nova peptide sequencing. *Jour­ nal of Computational Biology,* 16(8):1169-1182, 2009. ISSN 1066-5277. doi: 10.1089/cmb.2009.0122.
5. Andrew P. Horton, Scott A. Robotham, Joe R. Cannon, et al.

Comprehensive de nova peptide sequencing from ms/ms pairs generated through complementary collision induced dis­ sociation and 351 nm ultraviolet photodissociation. *Analytical Chemistry.* 89(6):3747-3753, 2017. ISSN 0003-2700. doi:

10.1021lacs.analchem.7b00130.

1. Andreas Bertsch, Andreas Leinenbach, Anton Pervukhin, et al. De nova peptide sequencing by tandem ms using complementary cid and electron transfer dissociation. *Elec­ trophoresis,* 30(21):3736-3747, 2009. ISSN 01730835. doi: 10.1002lelps.200900332.
2. Marie-Paule Lefranc and Gerard Lefranc. lmmunoglobulins or antibodies: lmgt® bridging genes, structures and func­ tions. *Biomedicines,* 8(9):319, 2020. ISSN 2227-9059. doi: 10.3390lbiomedicines8090319.
3. Carsten Schmelter, Natarajan Perumal, Sebastian Funke, et al. Peptides of the variable igg domain as poten­ tial biomarker candidates in primary open-angle glaucoma (poag). *Human Molecular Genetics,* 26(22):4451-4464, 2017. ISSN 0964-6906. doi: 10.1093/hmglddx332.
4. Vaibhav Singh, Marcel P. Stoop, Christoph Stingl, et al.

Cerebrospinal-fluid-derived immunoglobulin g of different **mul­** tiple sclerosis patients shares mutated sequences in comple­ mentarity determining regions. *Molecular and Cellular Pro­ teomics,* 12(12):3924-3934, 2013. ISSN 15359476. doi: 10.1074lmcp.m113.030346.

1. Ingrid Broodman, Dominique de Costa, Christoph Stingl, et al. Mass spectrometry analyses of K and *>,.* fractions re­ sult in increased number of complementarity-determining re­ gion identifications. *Proteomics,* 12(2):183-191, 2012. ISSN

16159853. doi: 10.1002lpmic.201100244.

1. 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.
2. 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.

1. Philip Lossl, Joost Snijder, and Albert J. R. Heck. Boundaries of mass resolution in native mass spectrometry. *Journal of the Amelican Society for Mass Spectrometry,* 25(6):906-917, 2014. ISSN 1044-0305. doi: 10.1007ls13361-014-0874-3.

**1**

1. Yuan Mao, Santosh G Valeja, Jason C Rouse, Christopher L Hendrickson, and Alan G Marshall. Top-down structural anal­ ysis 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: 10.1021lac303525n.

1. Yury O Tsybin, Luca Fornelli, Carsten Stoermer, et al. Struc­ tural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Analytical Chem­ istry,* 83(23):8919-8927, 2011. ISSN 0003-2700. doi: 10. 1021lac201293m.
2. Anja Resemann, Wolfgang Jabs, Anja Wiechmann, et al.

Full validation of therapeutic antibody sequences by middle­ up mass measurements and middle-down protein sequenc­ ing. *mAbs,* 8(2):318-330, 2016. ISSN 1942-0862. doi: 10.1080/19420862.2015.1128607.

1. 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 dissoci­ ation. *Journal of the Amelican Society for Mass Spec­ trometry,* 32(6):1326-1335, 2021. ISSN 1044-0305. doi:

10.1021/jasms.Oc00461.

1. Maurits A. den Boer, Jean-Francois Greisch, Sem Tamara, Al­ bert Bondi, and Albert J. R. Heck. Selectivity over coverage in de nova sequencing of iggs. *Chemical science,* 11(43):11886- 11896, 2020. ISSN 2041-6520. doi: 10.1039/d0sc03438j.
2. Jean-Francois Greisch, Maurits A. den Boer, Szu-Hsueh Lai, et al. Extending native top-down electron capture dissocia­ tion to mda immunoglobulin complexes provides useful se­ quence tags covering their critical variable complementarity­ determining regions. *Analytical Chemistry,* 93(48):16068- 16075, 2021. ISSN 0003-2700. doi: 10.1021/acs.analchem.

1c03740.

1. Christian Spoerry, Jana Seele, Peter Valentin-Weigand, Christoph G. Baums, and Ulrich von Pawel-Rammingen. Iden­ tification 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 00219258. doi: 10.1074/jbc.m115.711440.
2. Frederik Lermyte, Yury 0. Tsybin, Peter B. O'Connor, and

Joseph A. Loo. Top or middle? up or down? toward a stan­ dard lexicon for protein top-down and allied mass spectrom­ etry approaches. *Journal of the Amelican Society for Mass Spectrometry,* 30(7):1149-1157, 2019. ISSN 1044-0305. doi: 10.1007/s13361-019-02201-x.

1. Luca Fomelli, Eugen Damoc, Paul M. Thomas, et al. Anal­ ysis of intact monoclonal antibody igg1 by electron trans­ fer dissociation orbitrap ftms. *Molecular and Cellular Pro­ teomics,* 11(12):1758-1767, 2012. ISSN 15359476. doi: 10.1074/mcp.m112.019620.
2. Luca Fornelli, Daniel Ayoub, Konstantin Aizikov, et al. Top­ down analysis of immunoglobulin g isotypes 1 and 2 with elec­ tron transfer dissociation on a high-field orbitrap mass spec­ trometer. *Journal of proteomics,* 159:67-76, 2017. ISSN 1876-7737. doi: 10.1016/j.jprot.2017.02.013.
3. Luca Fornelli, Kristina Srzentic, Romain Huguet, et al. Ac­ curate 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.

1. Jared B. Shaw, Weijing Liu, V. Yury Vasilev, et al. Direct de­ termination 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.

1. Jared B. Shaw, Neha Malhan, V. Yury VasilEv, et al. Se­ quencing grade tandem mass spectrometry for top-down pro­ teomics using hybrid electron capture dissociation methods in a benchtop orbitrap mass spectrometer. *Analytical chem­ istry,* 90(18):10819-10827, 2018. ISSN 1520-6882. doi:

10.1021/acs.analchem.8b01901.

#### Chapter 1

1. Yufeng Shen, Nikola Tolic, Paul D Piehowski, et al. High­ resolution ultrahigh-pressure long column reversed-phase liq­ uid chromatography for top-down proteomics. *Journal of chro­ matography A,* 1498:99-110, 2017. ISSN 1873-3778. doi: 10.1016/j.chroma.2017.01.008.
2. Bin Ma, Kaizhong Zhang, Christopher Hendrie, et al. Peaks: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spec­ trometry,* 17(20):2337-2342, 2003. ISSN 0951-4198. doi: 10.1002/rcm.1196.
3. Kira Vyatkina, Si Wu, Lennard J. M. Dekker, et al. De novo sequencing of peptides from top-down tandem mass spec­ tra. *Journal of Proteome Research,* 14(11):4450-4462, 2015. ISSN 1535-3893. doi: 10.1021/pr501244v.
4. 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/
5. Stefano R. Bonissone, Thiago Lima, Katherine Harris, et al.

Serum proteomics expands on high-affinity antibodies in im­ munized rabbits than deep b-cell repertoire sequencing alone. *bioRxiv,* page 833871, 2020. ISSN 2692-8205. doi: 10.1101/

833871.

1. Pavlo Gilchuk, Adrian Guthals, Stefano R. Bonissone, et al.

Proteo-genomic analysis identifies two major sites of vulnera­ bility 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.

1. Seong Won Cha, Stefano Bonissone, Seungjin Na, Pavel A. Pevzner, and Vineet Bafna. The antibody repertoire of col­ orectal cancer. *Molecular and Cellular Proteomics,* 16(12): 2111-2124, 2017. ISSN 15359476. doi: 10.1074/mcp.ra117.

000397.

1. Yana Safonova, Stefano Bonissone, Eugene Kurpilyansky, et al. lgrepertoireconstructor: a novel algorithm for antibody

pr401300m. repertoire construction and immunoproteogenomics analysis.

1

1. Jason J. Lavinder, Andrew P. Horton, George Georgiou, and *Bioinformatics,* 31(12):53--61, 2015. ISSN 1367-4803. doi: Gregory C. Ippolito. Next-generation sequencing and protein 10.1093/bioinformatics/btv238.

mass spectrometry for the comprehensive analysis of human [97] Stefano R. Bonissone and Pavel A. Pevzner. lmmunoglob- cellular and serum antibody repertoires. *Current opinion in* ulin classification using the colored antibody graph. *Journal chemical biology,* 24:112-20, 2015. ISSN 1879-0402. doi: *of Computational Biology,* 23(6):483-494, 2016. ISSN 1066- 10.1016/j.cbpa.2014.11.007. 5277. doi: 10.1089/cmb.2016.0010.

1. Jason J. Lavinder, Yariv Wine, Claudia Giesecke, et al. Iden- [98] Lidong He, Lissa C. Anderson, David R. Barnidge, et al. Anal- tification and characterization of the constituent human serum ysis of monoclonal antibodies in human serum as a model antibodies elicited by vaccination. *Proceedings of the National* for clinical monoclonal gammopathy by use of 21 tesla ft-icr *Academy of Sciences,* 111(6):2259-2264, 2014. ISSN 0027- top-down and middle-down ms/ms. *Journal of the American* 8424. doi: 10.1073/pnas.1317793111. *Society for Mass Spectrometry,* 28(5):827-838, 2017. ISSN
2. Jiwon Lee, Daniel R. Boutz, Veronika Chromikova, et al. 1044-0305. doi: 10.1007/s13361-017-1602-6.

Molecular-level analysis of the serum antibody repertoire in [99] Lidong He, Lissa C. Anderson, David R. Barnidge, et al. Clas- young adults before and after seasonal influenza vaccination. sification of plasma cell disorders by 21 tesla fourier transform *Nature Medicine,* 22(12):1456-1464, 2016. ISSN 1546170. ion cyclotron resonance top-down and middle-down ms/ms doi: 10.1038/nm.4224. analysis of monoclonal immunoglobulin light chains in human

1. Jiwon Lee, Philipp Paparoditis, Andrew P. Horton, et al. Per- serum. *Analytical Chemistry,* 91(5):3263-3269, 2019. ISSN sistent antibody clonotypes dominate the serum response to 0003-2700. doi: 10.1021/acs.analchem.8b03294.

influenza over multiple years and repeated vaccinations. Ce// [100] John R. Mills, David R. Bamidge, and David L. Murray. Detect- *Host and Microbe,* 25(3):367-3765, 2019. ISSN 19313128. ing monoclonal immunoglobulins in human serum using mass doi: 10.1016/j.chom.2019.01.010. 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 amy­ loidosis. *Blood Cancer Journal,* 9(2):16, 2019. ISSN 2044-

5385. doi: 10.1038/s41408-019-0180-1.

[102] Mathieu Dupre, Magalie Duchateau, Rebecca Sternke­ Hoffmann, et al. De nova 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 autoanti­ body 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 Saye, Jared 0. Kafader,

et al. Next-generation serology by mass spectrometry: Read­ out of the sars-cov-2 antibody repertoire. *Journal of Pro­ teome Research,* 21(1):274-288, 2022. ISSN 15353907. doi: 10.1021/acs.jproteome.1c00882.

(105] Jared 0. 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 sub­ classes 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.

1. Albert Bandt, 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) re­ veals donor-specific clonal repertoires with high longitudinal stability. *Frontiers in Immunology,* 12:5190, 2021. ISSN 1664-

3224. doi: 10.3389/fimmu.2021.789748.

1. Noortje de Haan, David Falck, and Manfred Wuhrer. Monitor­ ing of immunoglobulin n- and o-glycosylation in health and dis­ ease. *Glycobiology,* 30(4):226-240, 2020. ISSN 1460-2423.

doi: 10.1093/glycob/cwz048.

1. Qingrong Yan, Maggie Huang, Michael J. Lewis, and Ping Hu.

Structure based prediction of asparagine deamidation propen­ sity in monoclonal antibodies. *mAbs,* 10(6):901-912, 2018. ISSN 1942-0862. doi: 10.1080/19420862.2018.1478646.

1. Beale Beyer, Manfred Schuster, Alois Jungbauer, and Nico Lingg. Microheterogeneity of recombinant antibodies: An­ alytics and functional impact. *Biotechnology Journal,* 13 (1):1700476, 2018. ISSN 18606768. doi: 10.1002/biot.

201700476.



