

# Perspective and Outlook, Summary,

**Samenvatting**

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

# Summary

**ANTIBODIES** are essential to adaptive immunity and represent one of the most polymorphic proteins found in the human body. This polymorphism provides

the incredible flexibility seen in adaptive immunity and makes antibodies a true "per­ sonalized proteome", unique to each individual and adapted to their needs. However, this diversity also makes studying antibodies difficult. The chapters in this thesis de­ tail efforts to develop generalizable data analysis strategies for studying secreted antibody repertoires using mass spectrometry.

In **Chapter 2,** we highlight the importance of having generalized tools to effectively analyze large and complex proteomic datasets. While crosslinking MS has emerged as an attractive method to probe protein interactions, the complexity of dealing with protein interactions rather than individual proteins has resulted in the production of large amounts of data, making processing difficult, especially for whole proteome ex­ periments. To address this issue, we have developed a tool that is interactive and facilitates analysis and visualization of these large datasets. The tool directly han­ dles the output of XlinkX for Proteome Discoverer but can also be used with output from other platforms through a user-controllable text-file importer. It comes equipped with a spectrum viewer and supports preprocessing of replicate datasets, enabling easy handling of large amounts of data. We have also integrated data from protein databases eggnog and uniprot, which enables integrated gene ontology enrichment analysis, grouping based on function, and mapping of known post-translational mod­ ification sites, domains, and secondary structures. Another feature is length-based validation of detected crosslinks by mapping the crosslinked peptides onto known structures of proteins or protein complexes. In situations where no structure is avail­ able, homologous structures can be used. In such cases, crosslinked peptides are aligned to the homologous sequence to obtain a confident placement of the linked residues before the distance between these residues is calculated using the 3D struc-

**6**

ture. Crosslinks between two proteins with known structures where no structure of the complex is available can also be directly submitted to DisVis for visualization and quantification of the information content of distance restraints.

In **Chapter 3,** we show how recent advancements in intact protein mass spectrom­ etry allow for the detection of single lgG1 molecules in human serum. This enabled the construction of personalized secretory lgG1 repertoires. Despite there being an enormous number of theoretically possible clones, the observed secreted antibody repertoires were relatively simple, with only several hundred clones dominating at any given time point. Moreover, while the majority of the clones in these profiles were stable over time, we observed large changes in the repertoires following a sepsis episode. We also demonstrated that a combination of peptide- and protein-centric mass spectrometry could be employed to sequence individual clones directly from the serum. The peptide-centric approach provided extensive coverage, while the protein-centric (fragmentation) approach provided sequence information that is in­ herently grouped per clone. The synergy between these techniques was used to sequence a single highly abundant clone from the sample of one of our donors.

**Chapter 4** showcases the potential of profiling data to compare and characterize individual donors within a group. We constructed SlgA1 profiles for a cohort of six lactating women who received a SARS-CoV-2 vaccination. The resulting profiles complement findings from earlier ELISA-based titer level analysis of these samples, where a biphasic rise in spike-specific lgA was found. Our observations indicate the emergence of a heterogeneous polyclonal population of between 100 and 200 novel clones in all donors after vaccination. This vaccine-induced population is dominated by a persistent population of clones that appear shortly after the initial vaccination and persist until at least five days after the second. However, we also detect a popu- lation of clones that emerge more than three days after the second vaccination was administered, in every donor. In-depth analysis of a strong responder, selected by

**6**

ELISA and confirmed by our data, reveals that the second rise in spike-specific lgA coincides with an abundant second dose-induced population, highlighting the diver­ gent clonal makeup of what initially seemed like a symmetrical biphasic response. Additionally, we observed several highly abundant clones appear and subsequently disappear from the secreted repertoire over the course of ~40 days, showing that highly abundant clones do not necessarily persist over time.

In **Chapter 5,** we built upon the proof of concept for endogenous antibody sequenc­ ing presented in Chapter 2 to create a more standardized and broadly applicable workflow for sequencing antibody chains in mixtures using a combination of peptide­ and protein-centric mass spectrometry. The proposed approach sequences a tar­ get chain in a modular, three-stage process based on germline domains. It starts with sequencing the framework regions, followed by complementarity determining re­ gions with flanking framework regions, and ultimately full chain sequences. Through integration of middle-down fragmentation, we could resolve ambiguity in *de nova* se­ quence predictions for the hypervariable complementarity determining regions. To achieve this, we filtered candidate sequences by comparing their theoretical mass to the gap between adjacent framework regions. We demonstrated the effectiveness of this approach by accurately sequencing a single targeted chain in a pure mono­ clonal antibody sample, an equimolar mixture of three monoclonal antibodies, and a polyclonal serum sample. This approach provides a broadly applicable workflow that could be used in future studies to sequence complex samples with high accuracy, as well as a step towards full automation of the process.

**6**

# Perspective and Outlook

Recent work by Wolf et al.(1) has demonstrated that an individual's antibody titers are a poor marker of the frequency of memory B cells generated following SARS­ CoV-2, seasonal influenza, or EBV infection. If assessing humoral immunity through polyclonal antibody titers does not reflect an individual's ability to mount an antibody response, then we need to find additional ways to determine an individual's level of protection. The research presented in this thesis describes a promising strategy to achieve this goal through proteomic antibody repertoire characterization using mass spectrometry. By monitoring proteomic antibody dynamics at an unprecedented res­ olution, we can gain new insights into humoral immune responses by uncovering when, how, and why specific antibodies are generated in response to physiologi­ cal events. Coupled with targeted proteomic sequencing of endogenous antibodies, this approach holds exciting potential for drug discovery as integration of these tech­ niques into antibody drug development pipelines could lead to significant advance­ ments in the field. Moreover, large scale profiling of endogenous secreted antibody repertoires may lead to the definition of immune signatures for use in disease risk assessment, diagnostic classification, or measuring treatment effectiveness.

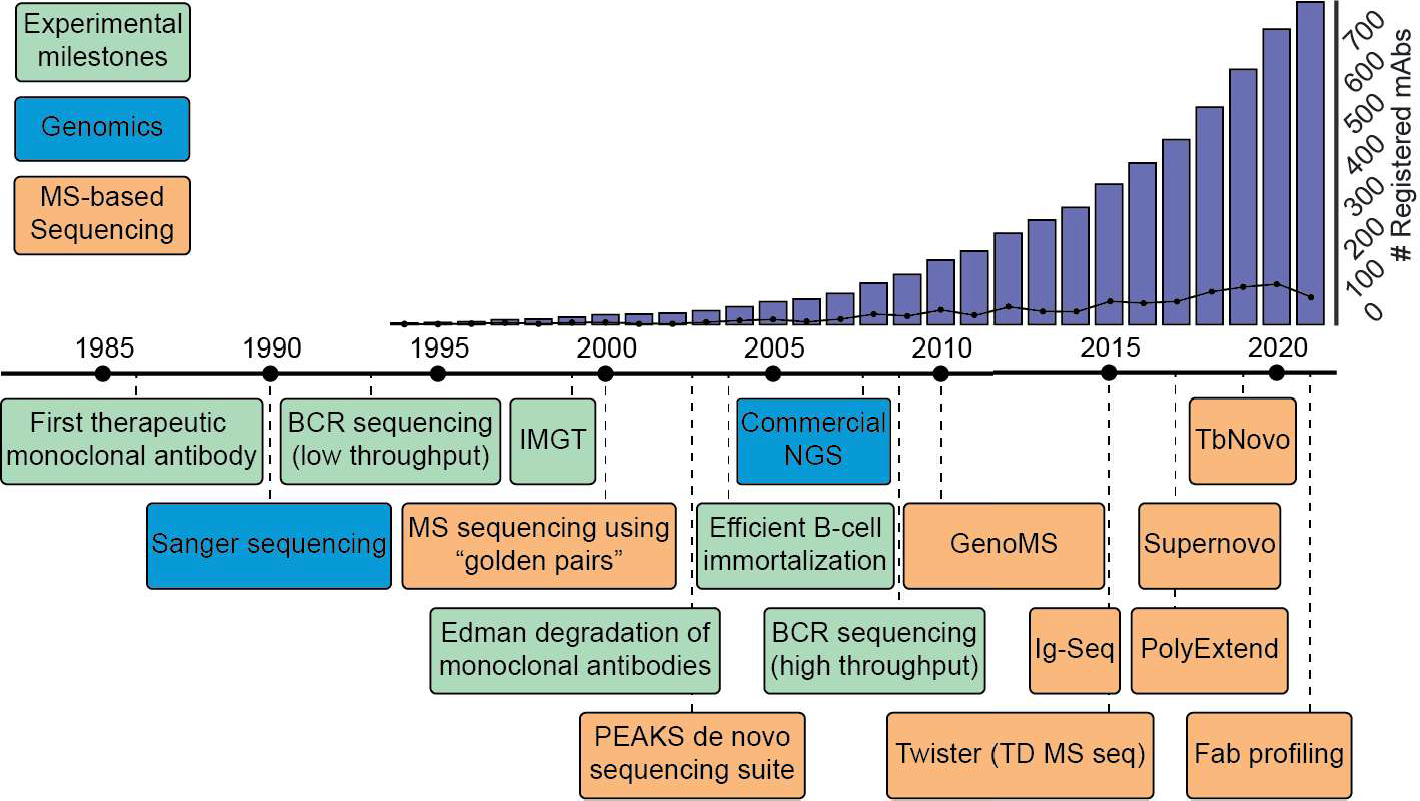
## The importance of standardized tools

**6**

The advancements made over the last decade in MS-based antibody sequencing provide an optimistic outlook for the future. I expect that a therapeutic antibody dis- covered by MS could be right around the corner. Looking back at the timeline of key developments in the field of antibody sequencing, we can notice several clear trends **(Figure 1**). Since the 1960s, rudimentary sample preparation for antibodies was available, but practical methods of obtaining sequence information appeared only in 1993, when Sanger sequencing was first applied to B cells. The first therapeutic an- tibody was registered in 1986, and this advent launched large-scale development of

mAbs, with a hundred mAbs registered by 2008 (2). At that point, next-generation sequencing led to high-throughput sequencing workflows and further facilitated the development of therapeutic antibodies. Over the last 20 years, the rapid expansion of genome-based sequencing techniques kick-started antibody discovery by allow­ ing large-scale BCR sequencing, and the number of deposited antibody sequences and registered antibody therapeutics has been growing exponentially ever since, with the 100th therapeutic mAb being approved by the United States Food and Drug Ad­ ministration in 2021 (3). Observing this trend, the popularization of MS-based pro­ teomics has now spurred the development of platforms for *de nova* sequencing of antibodies heavily supported by MS, and I envision that the ongoing advancement of MS based proteomic antibody profiling and *de nova* sequencing will complement available strategies by protein-level analysis. More specifically, monitoring of pro­ teomic antibody repertoires could be used to select a limited number of reactive antibodies from a patient with an effective immune response, which could then be sequenced, recombinantly produced and screened for neutralizing capacity. Such a direct approach to antibody discovery would be much more straightforward than screening of B-cells at the DNA/RNA level. The impact of accessible, standardized, high throughput analytical methods, can be observed in the discovery of genomic and proteomic biomarkers as well(4, 5). As standardized, high-throughput genomic and proteomic techniques became widely accessible, the number of genomic and proteomic biomarkers for disease risk assessment, early diagnosis, diagnostic clas­ sification and measuring treatment effectiveness skyrocketed. I believe the current advances in immune response characterization could represent a similar opportunity, as large-scale, in-depth characterization of proteomic antibody repertoires may lead to the discovery of defined immune signatures that could be used as immunological biomarkers in very similar ways.

**6**



**Figure 1: Timeline of key developments paving the way towards MS-based *de novo* sequencing of serum antibodies.** Blue: Key developments in the field of genomic sequencing. Green: Key advances in the field of antibody research. Orange: Selected hallmark papers in the field of MS-based antibody sequencing. To visualize the impact of therapeutic antibody development, the bar graph indicates the cumulative number of registered antibody-based drugs, and the line shows the number of registrations for a given year.(2)

## Challenges

**Larger sample size needed**

However, several challenges must be overcome before we can realize these goals. At the present time, there is simply not enough proteomic antibody repertoire data to draw generalizable conclusions about antibody repertoire dynamics. While we clearly observe drastic changes in the clonal abundance profiles in response to dis- ease and vaccination, the significance of these clones in relation to the antigen is not yet apparent. The fact that these repertoires are unique to each donor means we cannot simply compare the clonal repertoires of donors to screen for antibod- ies of interest, which, combined with the complex and heterogenous nature of these responses, makes finding patterns extremely challenging. Larger cohorts will need to be studied, and the obtained longitudinal antibody repertoires should be corre- lated to established techniques. Existing techniques like ELISA, BCR sequencing

**6**

and neutralization assays will be well complemented by the detailed characterization of the proteomic antibody repertoire. B cell receptor sequencing data could be used to determine the genetic and cellular origin of circulating clones, shedding light on whether novel clones are the result of somatic hypermutation or if they are the result of a completely new recombination of variable, diversity and joining (VDJ) gene re­ combination. Neutralization and binding assays could be used to determine exactly when an effective response emerges which can then be related to changes in the clonal profile. Such information on (individual) Fabs and the B-cells which produce them could be useful in studying the personalized nature of immune responses. As we increasingly correlate other assays to LC-MS Fab profiling data, we may also be able to distill a set of features common to effective neutralizing antibodies.

**Dealing with low-abundant clones**

A deep characterization of the proteomic antibody repertoire would also be highly desirable, as long lived, protective clones may not be among the most abundant fraction of the repertoire. While the observed proteomic repertoires were simple, consisting of 50-500 clones and dominated by a few abundant clones, the dynamic range of these secreted antibodies was wide and there may still be low abundant clones at concentrations below the current limit of detection. This is compounded by the fact that obtaining accurate uncharged, deisotoped masses for intact proteins (i.e. deconvolution) is an incredibly challenging task, particularly for low abundant species in complex samples. As we identify clones by mass and retention time, inconsistent mass determination impedes our ability to perform longitudinal tracking of clones in question and can lead to an underestimation of clonal longevity and an inability to deconvolute the LC-MS signal of a clone to an "antibody-like" mass (i.e., a mass of 45- 53 kDa) will result in failure to identify said clone. Similarly, a robust chromatography setup is required to prevent shifts in retention time due to an unstable system.

**6**

**Clonal lineage analysis and profiling ms**

Outside of experimental variation or artefacts of spectral processing, antibodies are highly polymorphic species and as such are subject to constant mutations which al­ most certainly lead to significant mass- and retention time shifts. Mutated clones would therefore show up as novel species in the antibody repertoire. While the ability to distinguish between these clones makes our analysis powerful, it also means that we are unable to identify clones with a shared clonal ancestor. While such clonal lineage analysis should in theory be greatly facilitated by referring detected masses to BCR-sequencing data, we have only observed a very small overlap between these two data streams. This apparent disjoint between the genomic or transcriptomic BCR sequences and proteomic data suggests that we are still missing a piece of the puz­ zle. A possible explanation lies in the sampling of the sequenced B cells. The most commonly used B cells for BCR sequencing are peripheral B cells, which only repre­ sent about 2% of the total B cell population (6).

**The need for sequencing/disjoint BCR**

While limited sequence information, in the form of sequence tags for example (7), can be used to reject the possibility of a shared clonal lineage, complete certainty requires full knowledge of the protein sequence of two clones. Unfortunately, in the current stage of implementation, *de nova* sequencing of antibodies is not suitable for such analysis or at least not at a significant scale, as it is still a highly challeng- ing task which requires manual curation by experts to derive the correct sequence. This manual curation is not only time-consuming, but also makes the process subject to interpretation errors when compared to more established sequencing techniques such as next generation DNA/RNA sequencing. As such, the need for automation is high, to improve not only the throughput but also the reproducibility and robustness.

6

**Sequencing-specific challenges**

Several challenges need to be overcome to achieve automated sequencing of abun­ dant clones in polyclonal mixtures. Similar to deconvolution of intact mass spectra (MS1), accurate and consistent deconvolution of fragmentation spectra (MS2) is im­ mensely challenging, doubly so as signal intensity is split across multiple fragments. Low abundant fragments are difficulty to acquire, and fragment coverage (i.e. the fraction of amide bonds that have one or more matching fragment mass in the spec­ trum) is typically limited (8). Peptide-centric analysis are hindered by the presence of other clones with homologous sequences, short peptide length and low depth of cov­ erage, which make read assembly exceptionally challenging. Nevertheless, leverag­ ing the synergy between the peptide- and protein-centric MS, as well as the available germline sequences has made sequencing of abundant clones in serum possible, and the generalizable workflow presented in this thesis can be used as a stepping­ stone towards full automation.

**Factors impacting sequencing strategies**

As antibody samples can be incredibly diverse, a big challenge for a robust antibody sequencing workflow is keeping the workflow broadly applicable. The optimal strat­ egy for a sequencing experiment depends on many factors: How many other clones are in the sample? Are these other clones relatively abundant compared to your target clone? Are specific proteases available to facilitate middle down analysis? How much sample is available? Are there coeluting clones? Is there the possibil­ ity for affinity purification or fractionation prior to sequencing? How divergent is the clone from the germline sequence? How good is bottom-up coverage? How good is top-down coverage? Many of these questions cannot be answered before starting the experiment or require additional experiments and therefore time, resources, and sample. While this may seem like a negative outlook, it is good to remember that just

**6**

a few years ago detection and deconvolution of individual antibody clones in com­ plex samples such as serum seemed practically impossible and sequencing even more so. The incredible advances in the field of biomolecular mass spectrometry over the past decades are an indication that there really is no telling how far we can still improve through incremental improvements, not only to spectral processing and acquisition but also to downstream processing of the data. In my opinion, the cur­ rent stage of implementation of *de nova* sequencing of endogenous antibodies has only scratched the surface of the possibilities, and there is a litany of opportunities to improve data analysis, instrumentation and protocol adaptations that are readily available.

**Protein centric improvements**

Nano-flow LC-MS can reduce sample requirements up to 100x, making it possible to acquire more middle down fragmentation data using the same sample. However, these systems are less robust than the high-flow systems used in our current ex­ perimental protocol (9, 10). Parallel acquisition strategies as seen in several 2D­ MS techniques could be used to boost signal intensity (11-13), and super resolution methods could provide greater resolving power (14). Improving precursor selection for fragmentation either by implementing some form of real-time processing (15) or

**6**

providing an inclusion list of target precursors based on a separate full MS spec- trum of the same sample. Additionally, while the current generalized implementation uses fragmented reduced antibody chains, fragmentation of intact fabs through ECO can yield highly complementary fragments which could be beneficial to the sequenc- ing process (16). Furthermore, additional fragmentation methods could be used to increase sequence coverage, as different fragmentation methods will preferentially fragment different backbone positions (17, 18). We have also landed on the Xtract and Respect deconvolution algorithms as our deconvolution method of choice, but there are alternatives out there. One drawback of our current methodology is that it

is restricted to deconvolution of scans from a single raw file and does not allow for manual selection of MS2 scans to average for deconvolution, instead only supporting averaging scans over a selected retention time window. Ms\_deisotope (19) is one such alternative that could be explored.

**Peptide centric improvements**

Our peptide centric efforts have been centered around using the *de nova* read as­ sembly tool stitch (20), combined with PEAKS *de nova* sequencing suite (21), both of which are under continued development along, as are *de nova* peptide sequenc­ ing algorithms and are thus likely to show improved performance over time. While the employed peptide-centric experimental strategies are highly optimized through selection of complementary proteases and defining a robust data acquisition method (22), we still find ways to improve throughput or performance regularly, for example through the use of SP3 beads (23), and potential future improvements could include providing exclusion lists of known constant region peptides.

## Conclusion

The work presented in this thesis emphasizes the need for advanced analytical strate­ gies in studying antibody dynamics and showcases the potential of MS-based pro- teomics as one such strategy. In doing so, I hope to have made studying these com­ plex systems less daunting for future researchers by outlining analytical strategies for untangling them. Historically, the complexity of antibodies has made them difficult to analyze on a large scale, as they lack the universal protein sequence databases used in other proteomic studies. However, recent advances in MS-based proteomics and *de nova* sequencing have paved the way for their inclusion, as evidenced by the in­ creasing number of published strategies for proteomic analysis of these repertoires. Combined with the immunological studies of unparalleled scale that we are seeing

**6**

since the SARS-CoV-2 pandemic, the stage now looks to be set for large scale, in depth, proteomic analysis of endogenous antibody repertoires. The incredible ad­ vances in studying these repertoires in multidisciplinary, collaborative studies has left me feeling highly optimistic about the future of this field. Through comprehensive analysis of endogenous antibody repertoires we can gain a deeper understanding of the immune responses they are a part of and I believe that mass spectrometry will play an integral role in untangling these personalized proteomes.

# Samenvatting

Antistoffen zijn essentieel voor adaptieve immuniteit en vertegenwoordigen een van de meest polymorfe eiwitten die in het menselijk lichaam warden gevonden. Deze polymorfie zorgt voor de ongelooflijke flexibiliteit die wordt gezien in adaptieve im­ muniteit en maakt antistoffen een ware "gepersonaliseerde proteoom", uniek voor elk individu en aangepast aan hun behoeften. Deze diversiteit maakt het echter oak moeilijk om antistoffen te bestuderen. De hoofdstukken in deze scriptie beschrijven inspanningen om algemeen toepasbare gegevensanalysestrategieen te ontwikkelen voor het bestuderen van uitgescheiden antilichaamrepertoires met behulp van mas­ saspectrometrie.

**6**

In **hoofdstuk 2** benadrukken we het belang van gegeneraliseerde tools om grate en complexe proteomische datasets effectief te analyseren. Hoewel crosslinking MS is opgekomen als een aantrekkelijke methode om eiwitinteracties te onderzoeken, heeft de complexiteit van het omgaan met eiwitinteracties in plaats van individuele eiwitten geleid tot de productie van grate hoeveelheden gegevens, waardoor de ver- werking moeilijk is, vooral voor experimenten met het hele proteoom. Om dit pro- bleem aan te pakken, hebben we een tool ontwikkeld die interactief is en analyse en visualisatie van deze grate datasets vergemakkelijkt. De tool behandelt rechtstreeks

de output van XlinkX voor Proteome Discoverer, maar kan ook worden gebruikt met output van andere platforms via een door de gebruiker te regelen tekstbestandsim­ porteur. Het is uitgerust met een spectra-viewer en ondersteunt voorverwerking van replicatiedatasets, waardoor het gemakkelijk is om grote hoeveelheden gegevens te verwerken. We hebben ook gegevens uit prote"fnedatabases zoals eggnog en uni­ prot ge"fntegreerd, wat ge"fntegreerde genontologie-verrijking, groepering op basis van functie en mapping van bekende posttranslationele modificatiesites, domeinen en secundaire structuren mogelijk maakt. Een andere functie is de lengtegebaseerde validatie van gedetecteerde crosslinks door de crosslinked-peptiden in kaart te bren­ gen op bekende structuren van eiwitten of eiwitcomplexen. In situaties waarin geen structuur beschikbaar is, kunnen homologe structuren worden gebruikt. In dergelijke gevallen worden crosslinked-peptiden uitgelijnd met de homologe sequentie om een vertrouwde plaatsing van de gekoppelde aminozuren te verkrijgen voordat de afstand tussen deze aminozuren wordt berekend met behulp van de 3D-structuur. Crosslinks tussen twee eiwitten met bekende structuren waarvan geen structuur van het com­ plex beschikbaar is, kunnen ook rechtstreeks aan DisVis worden aangeboden voor visualisatie en kwantificering van de informatie-inhoud van afstandsbeperkingen.

In **hoofdstuk 3** laten we zien hoe recente ontwikkelingen in intacte eiwitmassaspec­ trometrie de detectie van afzonderlijke lgG1-moleculen in menselijk serum mogelijk maken. Dit maakte de constructie van gepersonaliseerde secretoire lgG1-repertoires mogelijk. Ondanks het enorme aantal theoretisch mogelijke klonen, waren de waar­ genomen uitgescheiden antilichaamrepertoires relatief eenvoudig, met slechts en­ kele honderden klonen die op elk moment domineren. Bovendien waren de meeste klonen in deze profielen stabiel in de tijd, maar we observeerden grote veranderin­ gen in de repertoires na een sepsis-episode. We hebben ook aangetoond dat een combinatie van peptide- en eiwitgerichte massaspectrometrie kan worden gebruikt om individuele klonen rechtstreeks uit het serum te sequencen. De peptide-gerichte

**6**

aanpak gaf een uitgebreide dekking, terwijl de eiwitgerichte (fragmentatie) aanpak sequentie-informatie verschafte die inherent per kloon is gegroepeerd. De synergie tussen deze technieken werd gebruikt om een enkele sterk vertegenwoordigde kloon te sequencen uit het monster van een van onze donoren.

**Hoofdstuk 4** toont het potentieel van profilering gegevens om individuele donoren binnen een groep te vergelijken en te karakteriseren. We hebben SlgA1-profielen geconstrueerd voor een cohort van zes zogende vrouwen die een SARS-CoV-2- vaccinatie kregen. De resulterende profielen complementeren bevindingen uit eer­ dere ELISA-gebaseerde titeranalyse van deze monsters, waarbij een biphasische stijging van spike-specifieke lgA werd gevonden. Onze waarnemingen geven aan dat er na vaccinatie bij alle donoren een heterogene polyclonale populatie van tus­ sen de 100 en 200 nieuwe klonen ontstaat. Deze door het vaccin ge"induceerde populatie wordt gedomineerd door een aanhoudende populatie van klonen die kort na de eerste vaccinatie verschijnen en tot minstens vijf dagen na de tweede aanhou­ den. We detecteren echter ook een populatie van klonen die meer dan drie dagen na toediening van de tweede vaccinatie verschijnt, bij elke donor. Een diepgaande analyse van een sterke responder, geselecteerd door ELISA en bevestigd door onze gegevens, toont aan dat de tweede stijging van spike-specifieke lgA samenvalt met een overvloedige tweede dosis-ge"induceerde populatie, waarbij de divergente klo­ nale samenstelling van wat aanvankelijk leek op een symmetrische biphasische res­ pons wordt benadrukt. Bovendien hebben we waargenomen dat verschillende zeer overvloedige klonen verschijnen en vervolgens verdwijnen uit het uitgescheiden re­ pertoire gedurende ~40 dagen, wat aantoont dat zeer overvloedige klonen niet nood­ zakelijkerwijs aanhouden in de tijd.

6

In **hoofdstuk 5** bouwden we voort op het proof-of-concept voor endogene antilichaam­ sequentiebepaling dat in hoofdstuk 2 werd gepresenteerd, om een meer gestandaar­ diseerde en breed toepasbare workflow te creeren voor het sequencen van antili-

chaamketens in mengsels met behulp van een combinatie van peptide- en eiwit­ gecentreerde massaspectrometrie. De voorgestelde aanpak volgt een doelketen in een modulaire, drietrapsproces gebaseerd op kiemlijndomeinen. Het begint met het sequencen van de framework-regio's, gevolgd door de complementariteitsbepalende regio's met flankerende framework-regio's, en uiteindelijk volledige ketensequenties. Door integratie van middle-down fragmentatie konden we ambigu"fteit oplossen in de *de* novo-sequentievoorspellingen voor de hypervariabele complementariteitsbe­ palende regio's. Hiervoor filterden we kandidaat-sequenties door hun theoretische massa te vergelijken met de afstand tussen aangrenzende framework-regio's. We hebben de effectiviteit van deze aanpak aangetoond door een enkel gericht keten ac­ curaat te sequencen in een zuivere monoclonale antilichaammonster, een equimo­ lair mengsel van drie monoclonale antilichamen, en een polyclonaal serummonster. Deze aanpak biedt een breed toepasbare workflow die in toekomstige studies kan worden gebruikt om complexe monsters met hoge nauwkeurigheid te sequencen, en is daarnaast een stap richting volledige automatisering van het proces.

6

**References**

1. Carla Wolf, Sebastian Koppert, Noemi Becza, et al. Antibody levels poorly reflect on the frequency of memory b cells gen­ erated following sars-cov-2, seasonal influenza, or ebv infec­ tion. *Cells,* 11(22), 2022. ISSN 2073-4409. doi: 10.3390/

cells11223662.

1. 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: 10.1093/nar/gkz827.
2. Asher Mullard. Fda approves 100th monoclonal antibody prod­ uct. *Nature reviews. Drug discovery,* 20(7):491-495, 2021. ISSN 1474-1784. doi: 10.1038ld41573-021-00079-7.
3. Richard Simon. Genomic biomarkers in predictive medicine. an interim analysis. *EMBO Molecular Medicine,* 3(8):429, 2011. ISSN 17574676. doi: 10.1002/emmm.201100153.
4. Miao Cui, Chao Cheng, and Lanjing Zhang. High-throughput proteomics: a methodological mini-review. *Laboratory inves­ tigation;* a *journal of technical methods and pathology,* 102 (11):1170-1181, 2022. ISSN 1530-0307. doi: 10.1038/ s41374-022-00830-7.
5. Adrian Guthals, Yutian Gan, Laura Murray, et al. De nova
6. Christine L. Gatlin, Gerd R. Kleemann, Lara G. Hays, Andrew J. Link, and John R. Yates. Protein identification at the low femto­ mole level from silver-stained gels using a new fritless elec­ trospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Analytical Biochemistry,* 263 (1):93-101, 1998. ISSN 0003-2697. doi: 10.1006/abio.1998.

2809.

1. Catherine G. Vasilopoulou, Karolina Sulek, Andreas-David Brunner, et al. Trapped ion mobility spectrometry and pasef enable in-depth lipidomics from minimal sample amounts. *Na­ ture Communications,* 11(1):331, 2020. ISSN 2041-1723. doi:

10.1038ls41467-019-14044-x.

1. Mark E. Ridgeway, Markus Lubeck, Jan Jordens, Mattias Mann, and Melvin A. Park. Trapped ion mobility spectrometry: A short review. *International Journal of Mass Spectrometry,* 425:22-35, 2018. ISSN 1387-3806. doi: 10.1016/j.ijms.2018.

01.006.

1. Katherine A Graham, Charles F Lawlor, and Nicholas B Baratto. Characterizing the top-down sequencing of protein ions prior to mobility separation in a timstof. 2023. doi: 10.26434lchemrxiv-2022-w8s4n-v3.

ms/ms sequencing of native human antibodies. *Journal of Pro-* [14] Dmitry Grinfeld, Konstantin Aizikov, Arne Kreutzmann, Eugen

*teome Research,* 16(1):45-54, 2017. ISSN 1535-3893. doi: 10.1021/acs.jproteome.6b00608.

1. David L. Tabb, Ma Ze-Qiang, Daniel B. Martin, Amy Joan L. Ham, and Matthew C. Chambers. Directag: Accurate se­ quence tags from peptide ms/ms through statistical scoring. *Journal of Proteome Research,* 7(9):3838-3846, 2008. ISSN 15353893. doi: 10.1021/pr800154p.
2. Lidong He, Chad R. Weisbrod, and Alan G. Marshall. Protein de nova sequencing by top-down and middle-down ms/ms: Limi­ tations imposed by mass measurement accuracy and gaps in sequence coverage. *International Journal of Mass Spectrom­*

*etry,* 427:107-113, 2018. ISSN 13873806. doi: 10.1016/j.ijms.

2017.11.012.

1. Matthias Wilm and Matthias Mann. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry,* 68(1):1-8, 1996. ISSN 00032700. doi: 10.1021/ac9509519/asset/imagesl large/ac9509519f00006.jpeg.

Damoc, and Alexander Makarov. Phase-constrained spectrum deconvolution for fourier transform mass spectrometry. *Analyt­ ical Chemistry,* 89(2):1202-1211, 2017. ISSN 15206882. doi:

10.1021lacs.analchem.6b03636.

6

1. Kyowon Jeong, Masa Babovic, Vladimir Gorshkov, et al.

Flashida enables intelligent data acquisition for top-down pro­ teomics to boost proteoform identification counts. *Nature com­ munications,* 13(1), 2022. ISSN 2041-1723. doi: 10.10381 s41467-022-31922-z.

1. Maurits A. den Boer, Jean-Francois Greisch, Sem Tamara, **Al­** bert Bandt, 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: 10.1039ld0sc03438j.

(17] Mathieu Dupre, 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:

10.1021lacs.analchem.1c01955.

[18] 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.

(19] Joshua Klein, heckendorfc, Saulius Lukauskas, and mstim. mobiusklein/msdeisotope: Release v0.0.33. 2021. doi: 10. 5281/zenodo.5759830.

(20] Douwe Schulte, Weiwei Peng, and Joost Snijder. Template­ based assembly of proteomic short reads for de novo antibody sequencing and repertoire profiling. *Analytical chemistry,* 94 (29):10391-10399, 2022. ISSN 1520-6882. doi: 10.1021/acs.

analchem.2c01300.

(21] 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.

1. Weiwei Peng, Matti F. Pronker, and Joost Snijder. Mass spectrometry-based de novo sequencing of monoclonal an­ tibodies 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.
2. Harvey E. Johnston, Kranthikumar Yadav, Joanna M. Kirk­ patrick, et al. Solvent precipitation sp3 (sp4) enhances recov­ ery for proteomics sample preparation without magnetic beads. *Analytical Chemistry,* 94(29):10320--10328, 2022. ISSN

15206882. doi: 10.1021/acs.analchem.1c04200/supplfile/ ac1c04200si002.xlsx.



