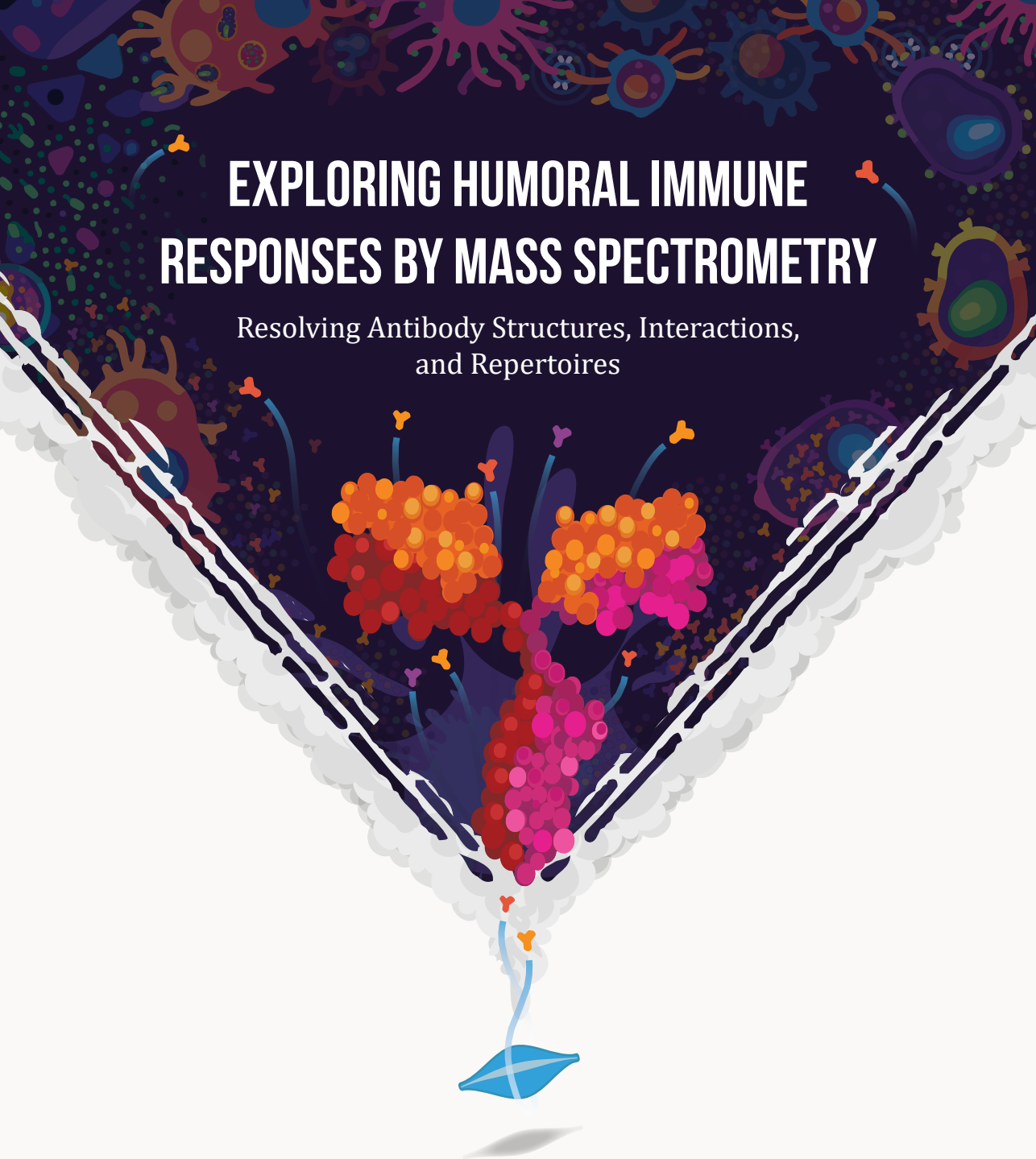
EXPLORING HUMORAL IMMUNE RESPONSES BY MASS SPECTROMETRY

Resolving Antibody Structures, Interactions, and Repertoires



MAURITS A. DEN BOER

“The best way to predict the future is to invent it.”

- Alan Kay

Exploring Humoral Immune Responses by Mass Spectrometry
PhD Thesis, Utrecht University, Utrecht, The Netherlands

Author: Maurits A. den Boer

ISBN: 978-94-6483-391-1

DOI: 10.33540/1942

Cover design and layout: Maurits den Boer, inspired by the book “Immune” by Philipp Dettmer

Printed by: Ridderprint | www.ridderprint.nl

The work described in this thesis was performed in the Biomolecular Mass Spectrometry and Proteomics group at Utrecht University in collaboration with the Department of Medical Microbiology at the University Medical Center Utrecht. Funding was provided through TTW-NACTAR Grant 16442 from the Netherlands Organization for Scientific Research (NWO) with contributions from Genmab B.V. and Hycult Biotechnology B.V.

Financial support for the printing costs of this thesis was kindly provided by Genmab B.V.

Exploring Humoral Immune Responses by Mass Spectrometry

Resolving Antibody Structures, Interactions, and Repertoires

De Humorale Immuunrespons Bestudeerd met Massaspectrometrie

Verkenning van Antilichaamstructuren, -interacties en -repertoires

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 9 oktober 2023 des middags te 2.15 uur

door

Maurits Arij den Boer

geboren op 5 juli 1991
te Raamsdonk

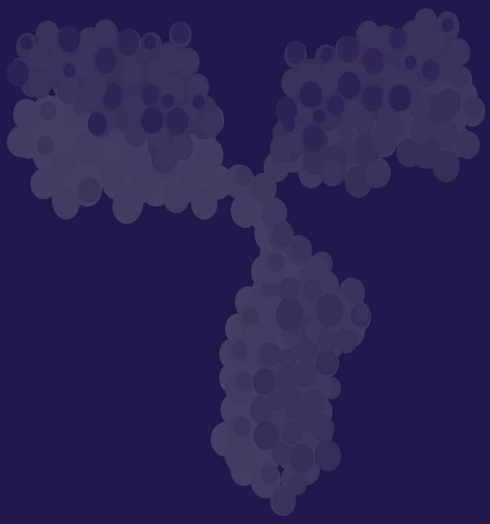
Promotoren:
Prof. dr. A.J.R. Heck
Prof. dr. S.H.M. Rooijakkers

Beoordelingscommissie:
Prof. dr. T. Logtenberg
Prof. dr. A.A. Makarov
Prof. dr. P.W.H.I. Parren
Dr. J. Schuurman
Prof. dr. G. Vidarsson

Table of Contents

PART 1

INVESTIGATING ANTIBODY BIOLOGY THROUGH MASS ANALYSIS



CHAPTER 1

INTRODUCTION TO ANTIBODY BIOLOGY AND ITS CHARACTERIZATION BY BIOMOLECULAR MASS SPECTROMETRY

In **Part 1** of my thesis, I describe the work that I performed to investigate the molecular details of antibody biology through mass analysis, using mass spectrometry and mass photometry as my primary tools. I will start with a concise introduction to antibodies and their role in immune protection against infections and other diseases. This section will cover the different isotypes and structures of human antibodies, as well as their interplay with other components of the immune system. I will continue with a short section to introduce biomolecular mass spectrometry (MS), which is expanded upon in an in-depth review of high-resolution native MS in **Chapter 2**.

1.1 Antibody-Mediated Immune Protection

Antibodies (Abs) are a fundamental component of our immune system's toolbox for battling harmful agents. Also called immunoglobulins (Igs), these proteins have a unique role in recognizing and neutralizing pathogenic substances, including bacteria, viruses, toxins and venoms[1]. When the body encounters such a threat, the adaptive immune system uses intricate mechanisms to generate antibodies that can detect and target the threat. This involves the presentation of foreign material to B lymphocytes, which develop effective antibodies in the iterative process of affinity maturation (see **Chapter 6**). Molecular structures capable of eliciting such responses are called antigens, on which antibodies are trained to recognize a distinctive molecular feature, an epitope. This epitope is bound by its complementary paratope on the resulting antibody, fitting together like a lock and key to ensure highly specific binding. Once effective antibodies have been generated, the body produces these in large quantities for release into the bloodstream and other bodily fluids. In this way, they patrol the body for their target, flagging it for the immune system to provide protection.

It should come as no surprise that the ability to harness the power of antibodies has become invaluable to both fundamental research and medicine[2,3]. Recombinantly produced monoclonal antibodies (mAbs) can be used to target molecular structures of interest with very high specificity, which enables their detection in biological research and diagnostics. Furthermore, they present powerful therapeutic agents that can effectively block bioactive molecules, neutralize viruses, or induce the clearance of pathogenic material, including cancers.

Antibody Structure

Structurally, antibodies are often represented as Y-shaped glycoproteins that bridge molecular recognition with the recruitment of immune functions (**Figure 1A**)[1,4]. Antigen recognition is facilitated by two identical Fragment antigen binding (Fab) arms that harbor the unique paratopes of the antibody. Effector functions on the other hand are mediated by the Fragment crystallizable (Fc) tail. The Fabs are typically connected to the Fc by a flexible hinge, which facilitates their orientational freedom. On the level of primary structure, antibodies are made up of four polypeptide chains that are linked by disulfide bonds. Two identical heavy chains (HC) of 50-70 kDa run along the full length of the Y-shape, with disulfide bonds connecting them in the hinge or Fc. Furthermore, each of the two identical Fab arms also harbors a light chain (LC) of about 23 kDa, connected to the HC through a single disulfide bond.

Both HC and LC are divided into multiple immunoglobulin domains of about 110 amino acids in length that each consist of two β-sheets. The most N-terminal domains are variable in sequence (variable heavy and variable light, VH and VL). They each contain three complementary-determining regions (CDRs) to shape the antigen-binding paratopes, with the CDR3 often said to be the most important[5,6]. The following domains are constant in sequence (CH_{1-4} and CL) and are shared between all antibodies of the same isotype. Especially the CH domains of the Fc interact with other proteins, such as other

antibodies and receptors of the immune system. These parts are also variably *N*-glycosylated, which is partly used to regulate these interactions, with some antibody isotypes holding additional *O*-glycosylation sites in the hinge between the CH1 and CH2 domains[7,8].

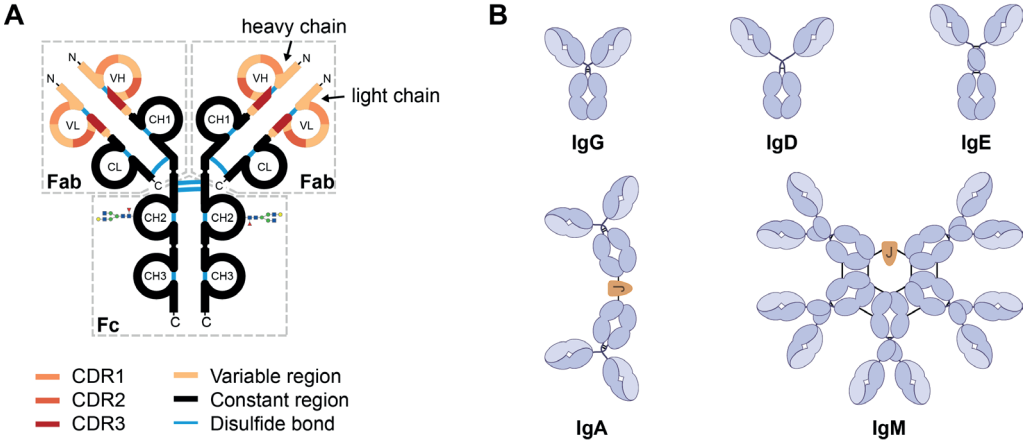


Figure 1 | Human antibody structures and classes. (A) Schematic overview of the primary structure of IgG1, the most abundant subclass in human circulation. The antibody Y-shape consists of two Fragment antigen binding (Fab) arms and one Fragment crystallizable (Fc) tail, which are connected by a flexible hinge. This structure is comprised of two identical heavy chains (HC) and two identical light chains (LC), which are interlinked by disulfide bonds (blue). Each chain starts with a variable domain (VH and VL) that harbors three complementary determining regions (CDRs) to make up the paratopes. The following domains are constant in sequence (CH1-3 and CL), with the CH2 domain containing an *N*-glycosylation site. **(B)** Overview of human antibody classes and their structures. IgA, IgG and IgD have a hinge that connects the two HCs by disulfide bonds and facilitates orientational flexibility of the Fabs. IgE and IgM on the other hand have an additional constant domain instead and are more rigid in structure. Both IgA and IgM can be produced in the presence of joining (J) chain, which leads to the formation of predominantly J-chain-linked IgA dimers and exclusively J-chain-linked IgM pentamers.

Antibody Classes

Human antibodies come in various isotypes – or classes – that differ in their effector functions and level of affinity maturation towards the antigen[1]. The isotype is determined by the HC, which in humans can be one of five forms: α, γ, δ, ε, and μ, corresponding to IgA, IgG, IgD, IgE and IgM antibodies (**Figure 1B**). All responses start with the generation of IgM, which serves as a first line of defense. In later stages, isotype switching and further optimization can occur to produce more mature antibodies of alternative isotypes.

While most antibodies are monomeric, some are present in one or more oligomeric states. Both IgA and IgM can be produced with a joining (J) chain, which facilitates the covalent linkage of additional monomers and secretion through the polymeric immunoglobulin receptor (pIgR). Although circulatory IgA is mostly monomeric, production in the presence of the J-chain leads to the formation of predominantly J-chain-linked dimers in secretory IgA. IgM on the other hand can be found as a J-chain linked pentamer under normal physiological conditions, both in circulatory and in secreted form. This

oligomerization is thought to improve antigen recognition through avidity effects, which is especially advantageous for relatively immature IgM responses.

IgG is the most abundant and most mature isotype in human circulation. This monomeric isotype is further subdivided into four subclasses (IgG1-4, in order of typical abundance in plasma) that have between 83% and 96% sequence similarity and differ in their effector functions. Differences are mostly in the hinge and just below in the CH2 domain, where these antibodies interact with other components of the immune system[9,10]. The number of disulfide bonds in the hinge also varies quite substantially between the subclasses, as does the location where the LC is linked to the HC[9]. Additionally, the IgG subclasses differ in the types of antigens that they recognize. IgG1 is thought to be active mostly against proteinaceous antigens, although IgG3 and IgG4 are also induced at lower abundance[11]. Polysaccharide antigens such as bacterial capsids and viral glycan structures, on the other hand, often induce IgG2 responses[11,12]. Because IgG1 is monomeric and relatively simple in terms of glycosylation and disulfide bond linkages, not surprisingly, this isotype is also the format of choice for antibody therapeutics.

Effector Functions and Complement Activation

After target recognition, there are several ways by which antibodies can elicit immunological effector mechanisms to provide protection[13]. The most straightforward of which is the neutralization of pathogenic material[14,15]. Neutralizing antibodies (nAbs) bind to bioactive molecules, shielding them from interaction with the host. This mode of action is particularly relevant for preventing viral entry into host cells, but bacterial toxins and other foreign molecules can also be counteracted in this way. Neutralization is fundamentally driven by strong affinity antigen binding, and often a single clone can already be enough to produce an effect.

Triggering more complex effector functions greatly benefits from avidity effects driven by immune complexation, the process by which multiple antibodies come together to recognize and bind a target[16]. Polyclonal mixtures are thought to be more effective in this respect, as multiple antibodies binding different antigens can more easily reach higher density and valency levels. When activation threshold levels are reached, these antibodies can interact with receptor proteins to trigger immune reactions. Fc receptor proteins (FcRs) binding IgGs (FcγRs) are particularly relevant for interaction with cytotoxic and phagocytic immune cells[17,18]. These cells can release cytotoxins to destroy infected host cells or can engulf pathogens for phagocytosis, the process by which they “eat” them. The neonatal Fc receptor (FcRn) in turn is an IgG receptor with functions in both immunity and transport[19]. FcRn binds IgGs in acidic environments, such as the internal endosomes phagocytes, and releases them at neutral pH. This enables the recycling of internalized IgGs back over the cell membrane and helps process the bound antigen for presentation on major histocompatibility complexes (MHCs). Secondly, FcRn is also found in many other cell types, where it plays an important role in IgG transport and regulation of IgG serum half-life.

Antibody clusters on antigenic surfaces can also recruit components of the humoral immune system (*i.e.*, reactions mediated by macromolecules), including the complement system[20-22]. This proteolytic cascade is comprised of over 30 plasma proteins that help recognize and clear pathogenic microorganisms and damaged host cells. There are three activation routes that differ in the way they recognize molecular patterns on cellular surfaces. For the antibody-dependent classical pathway, recognition is facilitated by IgM or IgG oligomers on the microbial surface that recruit complement component 1 (C1) (**Figure 2**)[23]. The multivalent antibody receptor of this complex, C1q, binds antibodies on their lower hinge and upper Fc through six interconnected globular heads. Binding to IgM is facilitated when naturally occurring oligomers bind multivalently to antigenic surfaces, where a conformational change exposes sequestered binding sites[24,25]. Monomeric IgG on the other hand requires non-covalent hexamerization for C1q recognition, which is facilitated by accumulation on antigenic surfaces[26-28]. When C1q binds to either of these surface-associated antibody clusters, its associated proteases C1r and C1s are activated. This in turn results in the cleavage of C4 and C2 to generate C4b2b, a C3 convertase that transduces the complement activation signal.

C3 conversion is considered a key process in the complement cascade, as it is the convergence point of the different activation pathways and fuels the amplification of the response[29]. C3 conversion on a bacterium has three main effects. Firstly, directly released C3a and downstream released C5a serve as chemoattractants to recruit immune cells. Simultaneously, C3b is covalently deposited on the bacterial surface, facilitating the recognition by complement receptors on phagocytes that engulf and digest them[30]. Lastly, C3 conversion can initiate the terminal complement pathway, which results in the formation of membrane attack complex (MAC) pores that can directly lyse Gram-negative bacteria[31,32]. Combined, this gives the complement cascade an important role in antibody-dependent immune protection against bacteria. Even more, it is also a crucial effector mechanism for therapeutic antibodies against cancer, which aim to induce a similar response to pathogenic host cells[33,34].

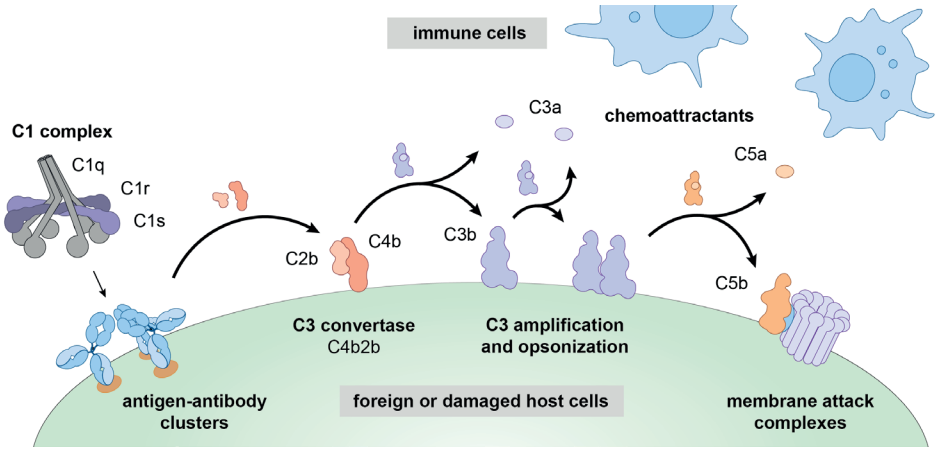


Figure 2 | Complement activation through the classical pathway. The classical pathway of complement is activated by IgM or IgG clusters on antigenic surfaces that recruit C1. When its constituent proteases C1r and C1s are activated by this process, they cleave C2 and C4 to form the C3 convertase C4bC2b (previously named C4b2b). Conversion of C3 transduces and amplifies the response, leading to the release of chemoattractants (C3a and C5) and coating of the surface by C3b. This allows for recognition by complement receptors on phagocytes, as well as the formation of membrane attack complexes (MAC).

The Potential of Recombinant Antibodies for Treating Infectious Diseases

While the strengths of antibody treatments for cancer and autoimmune diseases have been recognized for decades, more recently, infectious diseases have also come into focus. The field particularly received a boost during the COVID-19 pandemic, when neutralizing antibodies from recovered patients were quickly adapted for therapeutic use[35,36]. This demonstrated the viability of antibody therapies outside of their traditional use cases, highlighting their promise as a potential solution for the growing challenges in infectious diseases. Bacterial resistance to antibiotics for example is a quickly developing problem, which was estimated to be responsible for about five million annual deaths in 2019[37]. Particularly cases of *Escherichia coli*, *Staphylococcus aureus* and *Klebsiella pneumoniae* have become alarming in recent years, with the occurrence of more and more resistant strains[38-41]. Therapeutic antibodies that quickly and specifically direct the host immune system towards these pathogens could thus be particularly useful, especially if they can effectively activate the complement cascade[42-44]. Accordingly, when dealing with the challenges of future medicine, the importance of understanding antibody responses and complement activation cannot be overstated.

1.2 Biomolecular Mass Spectrometry for the Analysis of Antibodies and their Interactions

Biomolecular mass spectrometry (MS) has emerged as a powerful and versatile tool for the analysis of proteins and their interactions[45,46]. MS is an analytical technique that can extract a wealth of information from the (quantitative) measurement of the mass-to-charge ratio (*m/z*) of ions. The technology depends on the ionization of the analyte, which for biological molecules is often achieved through electrospray ionization (ESI)[47]. The ions are then routed into and through the (ultra) high vacuum of the mass analyzer, where they can be modulated to induce dissociation, fragmentation, or size-based separation for example. Next, the ions are transferred to a detector that records a mass spectrum containing the abundance and distribution of ions over the *m/z* range of the instrument.

To date, the most widespread form of MS-based analysis of proteins is bottom-up proteomics[45,46]. This approach analyses the proteome – the entirety of all proteins present – at the level of peptides resulting from a proteolytic digest of purified proteins, cell lysates, or even biopsies. The resulting complex peptide samples are typically separated by online reversed-phase (ultra) high-performance liquid chromatography ((U)HPLC) coupled to ESI-MS. In a two-layered MS approach, first, the intact masses of the peptides are recorded (MS1). Then, abundant precursor ions are mass-isolated and subjected to ion activation, followed by the recording of fragment ion spectra that provide information on the primary structure of the peptide (MS2). This combination of precursor and fragment masses is then searched against a protein or DNA database to assign them to the most probable matching protein or gene. Combined, this enables the identification of proteins in the sample, as well as the analysis of any modifications that are present. Moreover, using label-free quantification (LFQ) or intensity-based absolute quantification (IBAQ), these peptide MS data can also be used to quantify protein abundances.

On the other end of the biomolecular MS toolbox, native MS is a method that analyses intact proteins in a native-like folded state[48,49]. This is achieved by using volatile solutions with buffer-like characteristics such as aqueous ammonium acetate in combination with soft ESI-based ionization. With careful optimization, such experiments can maintain both tertiary and quaternary structures in the gas phase, allowing them to be studied[50]. This makes native MS very well-suited for the study of protein interactions, such as those between antibodies and other proteins of the immune system. Native MS and its applications to study intact proteins, proteoform profiles and protein complexes, is the subject of an extensive review I co-authored that is presented in **Chapter 2**.

1.3 Outline of Part 1

The chapters in **Part 1** of my thesis focus analytical technologies based on mass spectrometry and mass photometry, revealing how they can be used for the characterization of antibodies and their interactions.

In **Chapter 2**, I first continue the introduction with a review article I co-authored on high-resolution native mass spectrometry, an integral technology used in all scientific work described in this thesis. This chapter includes an extensive explanation of mass spectrometry instrumentation, as well as methods by which it can be used for the characterization of intact proteins, proteoforms, and protein-protein interactions.

The work presented in **Chapter 3** describes the benchmarking of mass photometry as a recently developed alternative technology for the mass analysis of proteins. Together with coworkers, I apply this technique for the characterization of antibodies and their interactions in heavily glycosylated immune complexes, comparing results between mass photometry and complementary techniques such as native mass spectrometry and size-exclusion chromatography hyphenated to multi-angle light scattering.

In **Chapter 4**, we combine native mass spectrometry with biochemical and biophysical techniques to study how Protein A of *Staphylococcus aureus* interferes with antibody-dependent complement activation. This provided new insights into how such bacteria can evade immune clearance. Our data also revealed a potential engineering approach that could be used in future applications of antibody therapy.

Lastly, **Chapter 5** combines multiple mass-based tools to challenge textbook knowledge of IgM, revealing CD5L as an additional protein component universally present in circulatory IgM.

References

1. Schroeder, H.W., Jr. & Cavacini, L. Structure and function of immunoglobulins. *J Allergy Clin Immunol* **125**, S41-52 (2010).
2. Lu, R.-M. et al. Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* **27**, 1 (2020).
3. Alkan, S.S. Monoclonal antibodies: the story of a discovery that revolutionized science and medicine. *Nature Reviews Immunology* **4**, 153-156 (2004).
4. Weaver, C. & Murphy, K. in Janeway's Immunobiology, Edn. 9th 139-172 (Garland Science, 2016).
5. Barrios, Y., Jirholt, P. & Ohlin, M. Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor. *J. Mol. Recognit.* **17**, 332-338 (2004).
6. Kuroda, D., Shirai, H., Kobori, M. & Nakamura, H. Structural classification of CDR-H3 revisited: A lesson in antibody modeling. *Proteins: Structure, Function, and Bioinformatics* **73**, 608-620 (2008).
7. Jennewein, M.F. & Alter, G. The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol.* **38**, 358-372 (2017).
8. de Haan, N., Falck, D. & Wuhrer, M. Monitoring of immunoglobulin N- and O-glycosylation in health and disease. *Glycobiology* **30**, 226-240 (2019).
9. Liu, H. & May, K. Disulfide bond structures of IgG molecules. *MAbs* **4**, 17-23 (2012).
10. Vidarsson, G., Dekkers, G. & Rispens, T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Front. Immunol.* **5** (2014).
11. Ferrante, A., Beard, L.J. & Feldman, R.G. IgG subclass distribution of antibodies to bacterial and viral antigens. *Pediatr. Infect. Dis. J.* **9**, S16-24 (1990).

12. Siber, G.R., Schur, P.H., Aisenberg, A.C., Weitzman, S.A. & Schiffman, G. Correlation between serum IgG-2 concentrations and the antibody response to bacterial polysaccharide antigens. *N. Engl. J. Med.* **303**, 178-182 (1980).
13. Lu, L.L., Suscovich, T.J., Fortune, S.M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nature Reviews Immunology* **18**, 46-61 (2018).
14. Klasse, P.J. Neutralization of Virus Infectivity by Antibodies: Old Problems in New Perspectives. *Advances in Biology* **2014**, 157895 (2014).
15. VanBlargan, L.A., Goo, L. & Pierson, T.C. Deconstructing the Antiviral Neutralizing-Antibody Response: Implications for Vaccine Development and Immunity. *Microbiology and Molecular Biology Reviews* **80**, 989-1010 (2016).
16. Oostindie, S.C., Lazar, G.A., Schuurman, J. & Parren, P.W.H.I. Avidity in antibody effector functions and biotherapeutic drug design. *Nature Reviews Drug Discovery* **21**, 715-735 (2022).
17. Mellor, J.D., Brown, M.P., Irving, H.R., Zalcberg, J.R. & Dobrovic, A. A critical review of the role of Fc gamma receptor polymorphisms in the response to monoclonal antibodies in cancer. *J. Hematol. Oncol.* **6**, 1 (2013).
18. Hayes, J.M., Wormald, M.R., Rudd, P.M. & Davey, G.P. Fc gamma receptors: glycobiology and therapeutic prospects. *J Inflamm Res* **9**, 209-219 (2016).
19. Pyzik, M. et al. The Neonatal Fc Receptor (FcRn): A Misnomer? *Front. Immunol.* **10** (2019).
20. Lu, L.L., Suscovich, T.J., Fortune, S.M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nat. Rev. Immunol.* **18**, 46-61 (2018).
21. Sarma, J.V. & Ward, P.A. The complement system. *Cell and Tissue Research* **343**, 227-235 (2011).
22. Mathern, D.R. & Heeger, P.S. Molecules Great and Small: The Complement System. *Clin. J. Am. Soc. Nephrol.* **10**, 1636-1650 (2015).
23. Gaboriaud, C. et al. Structure and activation of the C1 complex of complement: unraveling the puzzle. *Trends Immunol.* **25**, 368-373 (2004).
24. Sharp, T.H. et al. Insights into IgM-mediated complement activation based on in situ structures of IgM-C1-C4b. *Proc. Natl. Acad. Sci.* **116**, 11900-11905 (2019).
25. Czajkowsky, D.M. & Shao, Z. The human IgM pentamer is a mushroom-shaped molecule with a flexural bias. *Proc. Natl. Acad. Sci.* **106**, 14960-14965 (2009).
26. Diebolder, C.A. et al. Complement Is Activated by IgG Hexamers Assembled at the Cell Surface. *Science* **343**, 1260-1263 (2014).
27. Wang, G. et al. Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Mol. Cell* **63**, 135-145 (2016).
28. Strasser, J. et al. Unraveling the Macromolecular Pathways of IgG Oligomerization and Complement Activation on Antigenic Surfaces. *Nano Lett.* **19**, 4787-4796 (2019).
29. Ricklin, D., Reis, E.S., Mastellos, D.C., Gros, P. & Lambris, J.D. Complement component C3 – The "Swiss Army Knife" of innate immunity and host defense. *Immunol. Rev.* **274**, 33-58 (2016).
30. Rooijakkers, S.H.M. et al. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nat. Immunol.* **6**, 920-927 (2005).
31. Bayly-Jones, C., Bubeck, D. & Dunstone, M.A. The mystery behind membrane insertion: a review of the complement membrane attack complex. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160221 (2017).
32. Xie, C.B., Jane-Wit, D. & Pober, J.S. Complement Membrane Attack Complex: New Roles, Mechanisms of Action, and Therapeutic Targets. *The American Journal of Pathology* **190**, 1138-1150 (2020).
33. Lee, C.-H. et al. IgG Fc domains that bind C1q but not effector Fcγ receptors delineate the importance of complement-mediated effector functions. *Nat. Immunol.* **18**, 889-898 (2017).
34. de Jong, R.N. et al. A Novel Platform for the Potentiation of Therapeutic Antibodies Based on Antigen-Dependent Formation of IgG Hexamers at the Cell Surface. *PLoS Biol.* **14**, e1002344 (2016).
35. Jones, B.E. et al. The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates. *Sci. Transl. Med.* **13**, eabf1906 (2021).
36. Hurt, A.C. & Wheatley, A.K. Neutralizing Antibody Therapeutics for COVID-19. *Viruses* **13**, 628 (2021).
37. Murray, C.J.L. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629-655 (2022).
38. Lee, A.S. et al. Methicillin-resistant Staphylococcus aureus. *Nature Reviews Disease Primers* **4**, 18033 (2018).

39. Guo, Y., Song, G., Sun, M., Wang, J. & Wang, Y. Prevalence and Therapies of Antibiotic-Resistance in Staphylococcus aureus. *Frontiers in Cellular and Infection Microbiology* **10** (2020).
40. Falagas, M.E., Lourida, P., Poulikakos, P., Rafailidis, P.I. & Tansarli, G.S. Antibiotic Treatment of Infections Due to Carbapenem-Resistant Enterobacteriaceae: Systematic Evaluation of the Available Evidence. *Antimicrobial Agents and Chemotherapy* **58**, 654-663 (2014).
41. Ballén, V. et al. Antibiotic Resistance and Virulence Profiles of Klebsiella pneumoniae Strains Isolated From Different Clinical Sources. *Frontiers in Cellular and Infection Microbiology* **11** (2021).
42. Oleksiewicz, M.B., Nagy, G. & Nagy, E. Anti-bacterial monoclonal antibodies: Back to the future? *Archives of Biochemistry and Biophysics* **526**, 124-131 (2012).
43. Motley, M.P., Banerjee, K. & Fries, B.C. Monoclonal antibody-based therapies for bacterial infections. *Curr. Opin. Infect. Dis.* **32**, 210-216 (2019).
44. Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J.D. Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* **11**, 785-797 (2010).
45. Altelaar, A.F.M., Munoz, J. & Heck, A.J.R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* **14**, 35-48 (2013).
46. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).
47. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246**, 64-71 (1989).
48. Heuvel, R.H.H.v.d. & Heck, A.J.R. Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8**, 519-526 (2004).
49. Leney, A.C. & Heck, A.J.R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass. Spectrom.* **28**, 5-13 (2017).
50. Bakhtiari, M. & Konermann, L. Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *The Journal of Physical Chemistry B* **123**, 1784-1796 (2019).

CHAPTER 2

HIGH-RESOLUTION NATIVE MASS SPECTROMETRY

Sem Tamara[1,2,*], Maurits A. den Boer[1,2,*], and Albert J.R. Heck[1,2]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

[2] *Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands*

[*] *Contributed equally*

Published as:

High-Resolution Native Mass Spectrometry
Sem Tamara, Maurits A. den Boer, and Albert J.R. Heck
Chemical Reviews 2022 122 (8), 7269-7326
DOI: 10.1021/acs.chemrev.1c00212

Abstract

Native mass spectrometry (MS) involves the analysis and characterization of macromolecules, predominantly intact proteins and protein complexes, whereby as much as possible the native structural features of the analytes are retained. As such, native MS enables the study of secondary, tertiary, and even quaternary structures of proteins and other biomolecules. Native MS represents a relatively recent addition to the analytical toolbox of mass spectrometry and has over the last decade experienced immense growth, especially in enhancing sensitivity and resolving power, but also in ease-of-use. With the advent of dedicated mass analyzers, sample preparation and separation approaches, targeted fragmentation techniques, and software solutions, the number of practitioners and novel applications have risen both in academia and industry. This review focuses on recent developments in high-resolution native MS, describing applications in the structural analysis of protein assemblies, proteoform profiling of – among others – biopharmaceuticals and plasma proteins, and quantitative and qualitative analysis of protein-ligand interactions, with the latter covering lipid, drug, and carbohydrate molecules, to name a few.

Contents

1. Introduction and Scope of This Review

Biological mass spectrometry comes in many flavors. This diversity originates not only from the many different biomolecules that can be analyzed and investigated, such as proteins, peptides, lipids, DNA, RNA, carbohydrates, and metabolites but also from the wide assortment of tools available to characterize them. Focusing on proteins and peptides, to date the dominant portion of MS-based analysis is performed by peptide-centric proteomics. Herein proteins are identified and quantified following enzymatic digestion into easily amenable smaller peptides, whose sequences can be determined by different fragmentation methods and matched by well-developed search algorithms against protein, RNA, and DNA databases. MS-based proteomics provides a means to measure proteome-wide protein abundances and monitor them upon perturbation of a system. Additionally, it can also be used to chart proteome-wide protein-protein interactions and various post-translational modifications (PTMs). Several excellent reviews are available covering all these distinct flavors of MS-based proteomics[1,2].

Beyond MS-based proteomics, mass spectrometry's role in biochemistry and biology has over the past decades expanded to cover many aspects of structural and molecular biology[3]. For MS-based structural biology – as is the case for conventional proteomics – quite a few different approaches contribute highly complementary information. Surface labeling techniques, involving either hydrogen-deuterium exchange[4,5] or chemical labeling with radicals or other small molecules[6], coupled with mass spectrometry can provide information about structural changes and interaction interfaces. Chemical cross-linking approaches coupled to mass spectrometry facilitate structural investigation by providing distance restraints between identified cross-linked amino acids and offer information about new protein-protein and even protein-DNA interactions[7-9]. These technologies still primarily use peptide-centric proteomics approaches and involve a proteolytic digestion step prior to LC-MS analysis.

In contrast to the technologies outlined above, native mass spectrometry (native MS), the core focus of this review, analyzes intact proteins and their non-covalent complexes, as well as other biomolecules, in a native-like folded state. Although the birth of native MS can be traced back to the early 1990s[10,11], just a few years after the introduction of electrospray ionization mass spectrometry (ESI-MS)[12], MS-based technologies involving the analysis of intact proteins and protein complexes are still not as mature as their peptide-centric counterparts, mainly because of the challenges behind efficient ionization and detection of the larger intact protein ions. In native MS, the aim is to bring the analyte into the mass analyzer while retaining its original native structure and inter- and intramolecular interactions as much as possible[13-17]. This task is not trivial, as the biomolecule is charged in the ionization process and stripped of all solvent molecules before mass analysis can occur under (ultra)high-vacuum conditions. Thus, a fully native state can never be retained. However, through a plethora of experimental work over the past decades, it has become apparent that when conditions are carefully managed, electrospray ionization may provide gas-phase ions of proteins and protein complexes that retain many of their native features[18,19]. The first examples of that focused on the analysis of intact non-covalent complexes, which remained intact and largely retained their quaternary structures throughout their transfer from the solvent into the gas phase

up until they hit the mass analyzer's detector. In the early years of this century, the term "native MS" was coined to describe this area of biomolecular mass spectrometry[14]. Native MS has matured substantially since then, and now many groups are applying this technology to study all sorts of proteins, their assemblies, and the interactions between proteins and ligands, including small molecule drugs, cofactors, lipids, nucleotides, DNA, and RNA (**Figure 1**). Although native MS was initially primarily used to study soluble protein assemblies, membrane protein complexes have recently entered the realm of native MS through electrospraying of these assemblies from detergent micelles, nanodiscs, or even native lipid membranes[20,21].

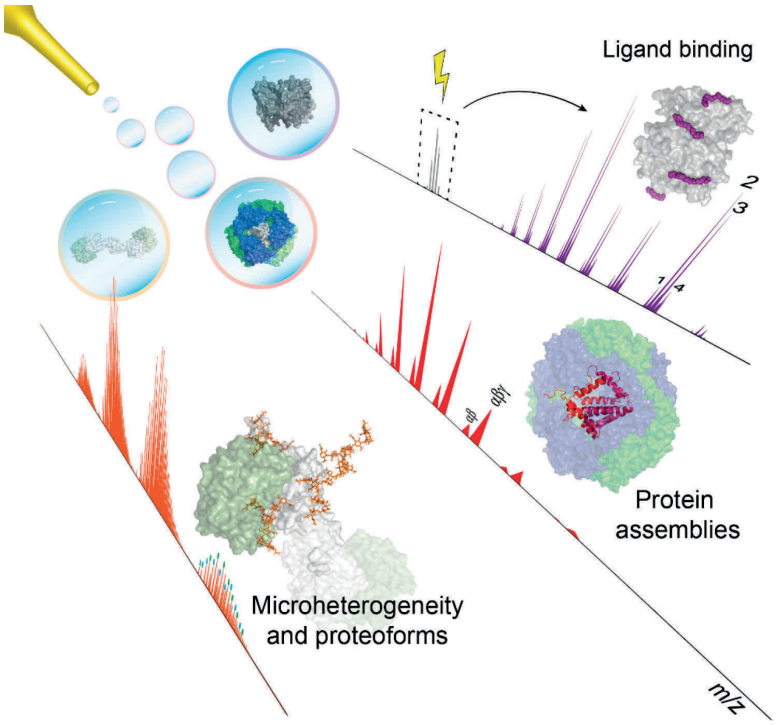


Figure 1 | Native MS has over the past decade experienced immense advances, primarily in enhancement of sensitivity and resolving power of mass analyzers but also in its ease of use. These advances have enabled the use of high-resolution native MS to analyze and characterize a wide range of macromolecules, including protein and ribonucleoprotein assemblies (**section 7**) and proteoforms of intact biopharmaceuticals and plasma proteins (**section 8**), and binding between proteins and small-molecule ligands such as lipids, drugs, and carbohydrates (**section 9**).

Significant advances have been made in native MS over the past decades, yet some analytical challenges remain to be overcome. Although we will focus on instrumentation in this review (**section 3**), one of the most critical challenges in native MS lies in sample preparation or, more specifically, making the analytes amenable to native electrospray ionization. This process is quite sample-dependent, and therefore, no universal guidelines can suffice. Nevertheless, it is crucial to recognize the important role of volatile salt solutions, such as ammonium acetate, that not only create a "native-like" environment for

biomolecules but also minimize adduct formation and enable facile solvent removal at low activation energies, consequently resulting in clean native spectra[22]. Although the principles of electrospray ionization are long known and have been discussed in detail[23,24], the mechanism of electrospray ionization of biomolecules has remained relatively elusive[25]. Alternative ionization methods and designs are also emerging, with desorption electrospray ionization being a notable example, enabling ionization of biomolecules directly from biological tissues[26].

Key instrumentational challenges today relate to the growing size and mass heterogeneity of the protein assemblies analyzed. The field of native MS is moving toward analyzing larger and larger protein assemblies, some with masses of several megadaltons or more. Moreover, increasing (micro)heterogeneity resulting from a plethora of PTMs (*e.g.*, phosphorylation, acetylation, glycosylation, lipidation) is introduced as the field moves away from bacterially produced recombinant proteins toward more endogenous analytes of eukaryotic origin. These shifts in the nature of the samples analyzed pose new challenges to the instruments used for native MS.

Large protein assemblies ionized by native electrospray attain relatively fewer charges than those ionized under denaturing conditions, thus requiring analysis using mass analyzers with extended m/z ranges (up to m/z = 20,000 Th and beyond). As a result, in the first decades, the field of native MS was highly dependent on time of flight (ToF)-based mass analyzers that have a theoretically unlimited m/z range. However, most standard mass analyzers are optimized for the transmission of relatively small molecules (*e.g.*, peptides), and thus, ToF-based instruments also needed to be modified for the dedicated mass analysis of protein complexes. Besides ToF-based mass analyzers, ion-trap-based instruments have also been modified and optimized for native MS, with developments on both Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap mass analyzers[27-30]. The development of commercial Orbitrap mass analyzers with extended mass ranges has given a strong impetus to the field of native MS, whereby next to academic laboratories, pharmaceutical and biotechnological companies have adopted the technology[30,31].

The other challenge, which is inherent to almost any mass spectrometric approach, is the mass resolving power, a central subject in **section 2**: the higher the mass resolving power, the more precise the mass analysis. This is as true for native MS as for any other form of mass spectrometry. However, the attainable mass resolving power in native MS is not only dependent on the instrumental mass resolution limits but also heavily affected by the ionization process and especially by the ability (or the lack thereof) to completely desolvate the protein or protein complex, stripping off all solvent molecules and salt adducts during the transfer from the electrospray solution into the (ultra)high vacuum of the mass analyzer[32]. Only when all these small molecules are removed can the mass and m/z of the detected ions correspond to the analyte's exact mass. For large proteins and protein complexes, such complete desolvation upon electrospray ionization is not trivial. Even if all of the water molecules can be removed, salt cations (*e.g.*, Na^+ or K^+) may remain attached because of their stronger binding, leading to ion signals that are broadened because they originate not only from the multiply protonated analytes but also from

analytes carrying metal cations, which slightly increase the mass and m/z of the ions. The presence of these small molecule adducts depends heavily on sample preparation, and extensive desalting is often a prerequisite for obtaining highly resolved mass spectra. However, this often comes at the expense of sample loss. On the other hand, desolvation can be promoted during the ionization process and while the ions are being transferred through the mass analyzer. This is often achieved via ion activation, either by heating the ion source or by letting the ions collide with inert gas molecules in a balanced manner, whereby the analyte ions lose their attached solvent adducts but do not undergo dissociation.

This review focuses on high-resolution native MS. In **section 3** we describe how technological innovations in mass analyzers have advanced native MS by extending the achievable mass (or m/z range), enhancing the sensitivity by optimizing the ion transmission, and improving the achievable mass resolving power. In particular, we focus on the latter and describe how different factors are optimized to enhance the mass resolving power, including improved desolvation by ion activation. In **section 6**, we describe different methods of tandem mass spectrometry (*i.e.*, collision-, electron-, and photon-induced) used in native MS to obtain more structural information on the analyzed protein complexes but also to improve desolvation and the precision of the mass measurement. **Sections 4** and **5** focus on parallel developments in advancing spectral data processing for native MS, which are geared toward the more efficient and improved analysis of the native mass spectra. Automated data processing becomes of the utmost importance when the analytes are very heterogeneous in mass and when the resolution in native mass spectra is hampered by imperfect desolvation or binding of various small-molecule cofactors.

In the remainder of the review, we describe various studies highlighting the recent advances and diverse applications of high-resolution native MS. Broadly, this part of the review is separated into several sections. **Section 7** includes an overview of studies on protein complexes by high-resolution native MS, wherein the generic aim is to determine the stoichiometry, structural features, subcomplexes, and possible binding of functionally important ligands (*e.g.*, lipids, DNA, and glycans). Next, in **section 8** we describe studies using high-resolution native MS to characterize intact single proteins, whereby the aim is to assess qualitatively and quantitatively all of the proteoforms present. Here we primarily focus on the analyses of protein biotherapeutics and plasma glycoproteins. In **section 9** we highlight progress in the field of characterizing protein-small molecule interactions by high-resolution native MS. We end the review with a future outlook.

2. Critical Measures of Performance in High-Resolution Native MS

With the ever-increasing size and complexity of the analytes, native MS needs to continually improve its performance regarding critical metrics such as mass accuracy, detection sensitivity, upper mass boundary, tandem MS capabilities, and mass resolution. High mass accuracy and sensitivity are essential for quality control analyses of large therapeutic molecules produced by the biopharmaceutical industry, as described in more detail below in **section 8**[33]. Mass range boundaries for native MS are constantly pushed by the desire to investigate larger macromolecules, such as viral particles and nanocontainers with molecular weights (MWs) in the megadalton mass range[34]. Finally, aside from detecting distinct charge states of large ionized macromolecular assemblies, mass resolution in native MS must distinguish between species with minute mass differences that stem from salt adducts, small non-covalently bound ligands, or PTMs. Such demands are continuously tackled by the research community and have led to the modern age of high-resolution native MS with the emergence of exciting new technologies and novel instruments, as further described below.

2.1. A Closer Look at Resolution and Resolving Power

Although it is an essential measure of performance in mass spectrometry, *mass resolution* is still often ill-defined as a term by parts of the community. According to the International Union of Pure and Applied Chemistry (IUPAC), *mass resolution* is defined as $m/\Delta m$, where m is typically the mass-to-charge ratio of a singly charged ion and Δm is determined either as the width of a single peak at a fraction of its height (*e.g.*, 50%) or as a mass difference between two equally abundant peaks with a valley between them not exceeding 10% of their heights[35]. On the other hand, *resolving power* is the ability of an instrument to distinguish between two peaks differing by a small m/z value and is defined as the peak width, Δm[36]. The misperception quite often observed in the literature arises because *mass resolving power* is defined similarly to the mass resolution as $m/\Delta m$[35]. The similarity between the definitions of these two terms has resulted in their somewhat interchangeable usage in the literature. In addition, different definitions of resolution and resolving power have been proposed[37-39], although they are not widespread. In the literature on mass spectrometric instrumentation, *mass resolution* is often defined as $m/\Delta m$, where Δm is determined at 50% peak height, also called the full width at half-maximum (fwhm)[40-42]. Here we adhere to defining both the mass resolution and mass resolving power using the fwhm since it is the most adopted definition in the native MS literature and is advised in the IUPAC terminology recommendations for mass spectrometrists[36]. While we use the term mass resolution throughout this review, we use the term mass resolving power to characterize the mass analyzer performance, as the latter is very commonly used in such context. It should be noted that it is important to consider the mass value at which the resolution or resolving power is defined, as – for many instruments, most notably in FT-ICR and Orbitrap instruments – the resolution does not scale up well with increasing m/z.

2.2. High Resolution in the Context of Native MS

In the field of small molecule and peptide-centric mass spectrometry, high resolution is primarily used for the disentanglement of isotope patterns, enabling the determination of accurate masses and – in the case of small molecules – even exact chemical formulas with sub-ppm mass accuracies achieved on modern FT MS instruments[43-45]. The absolute highest resolution in the range of several million (at m/z = 100-200 Th) was achieved with FT MS, especially FT-ICR, reporting resolutions far beyond 10^6 at m/z = 200 Th. This is closely followed by another representative of FT MS, the Orbitrap mass analyzer, which, although featuring lower resolution at low m/z, still surpasses 10^6 when measured at m/z = 200 Th. With increasing acceptance, such high-resolution measurements are referred to as ultra-high resolution and have led to significant advancements in the fields of petroleomics[46], proteomics[47], environmental analysis[48], forensics[49], and space exploration[50]. However, such ultra-high resolution is not necessarily beneficial for native MS, as outlined below.

Although measuring isotopically resolved spectra seems beneficial, above certain molecular weights (~ 150 kDa) the isotopic distributions of various codetected species, *e.g.*, ions carrying salt adducts or additional solvent molecules, start to overlap and superimpose, leading to distortions of the acquired signal (**Figure 2A**). This effect is increasingly more pronounced for larger molecules, whereby the detected peaks of globular proteins are substantially broader than expected (**Figure 2B**). In 2014 Lössl *et al*. argued that mass-resolving small buffer molecules or salt adducts is impossible for proteins larger than 65 kDa[32]. Moreover, in the case of large multiply charged biomolecules, PTM-related microheterogeneities result in only minute m/z differences, *e.g.*, oxidation of a 1 MDa molecule carrying 80 charges would lead to a shift of only m/z = 0.2 Th, while the width of the isotopic distribution is above 0.5 Th[34]. The current methods for desalting and complete desolvation of ions of large macromolecular assemblies are only partially effective for analytes in the megadalton range. Furthermore, the technical requirements for achieving sufficient resolution are often incompatible with large macromolecules. For example, in Orbitrap MS, the resolution is inversely proportional to the square root of m/z, requiring a stable image current signal to be recorded in the range of several seconds to gain isotopic resolution for larger macromolecular assemblies. These limitations hamper the analysis of large multiply charged protein assemblies with conventional native MS, whereby they are more prone to suffer from space-charge effects or decay due to field imperfections. As a result, in native MS, most researchers aim to resolve distinct charge states rather than individual isotope peaks while still endeavoring to disentangle microheterogeneities introduced by PTMs[51] and co-occurring stoichiometries[52].

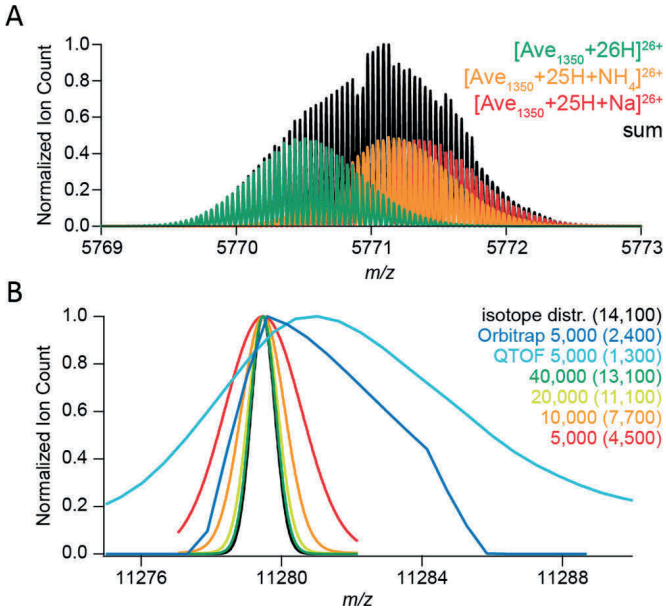


Figure 2 | Challenges for resolving isotopologues with high-resolution native MS. (A) Adduct ions affect the mass resolving power. Baseline isotope mass resolution does not permit bare, sodium-bound, and ammonium-bound ions of a 150 kDa averagine protein to be distinguished using native ESI-MS. According to the empirical charging behavior of globular proteins in native MS, 26+ is the most abundant charge state for a molecular weight of 150 kDa. Therefore, the peaks of 26-fold-charged cations of a 150 kDa averagine protein (1350 averagine residues) were generated with MassLynx version 4.1, assuming baseline isotope mass resolution (R = 500,000). The isotope distributions of unmodified (green), ammonium-bound (orange), and sodium-bound (red) protein ions were simulated individually and subsequently summed to produce their combined mass spectrum (black). **(B)** Experimental peaks of globular protein complexes are substantially broader than simulated peaks of their molecular ions. The apparent mass resolution depends on the preset instrumental resolution and the efficiency of adduct removal. Shown are mass spectra of a GroEL ion with a charge of 71+. These were measured on Orbitrap Exactive Plus (blue) and a QToF (cyan) instruments, both operating at an instrument mass resolution of 5000, or simulated with MassLynx version 4.1 at mass resolutions of 5000 (red), 10,000 (orange), 20,000 (yellow), and 40,000 (green). The black curve represents the natural isotope envelope of GroEL. Numbers in parentheses correspond to the apparent mass resolutions R_{nat} determined by measuring the experimental peak widths. Reproduced from ref [32]. Copyright 2014 American Society for Mass Spectrometry.

Similar to how narrow charge state distributions in native MS improve sensitivity, recording the average mass of an isotopic envelope instead of all of the peaks within the fine isotope distribution can provide further gains in sensitivity and signal-to-noise ratio (S/N), as the signal is binned over fewer channels. As a bonus, in conventional FT-based native MS, the signals of large ions decay less within the shorter acquisition durations provided by intermediate instrumental resolutions[53]. Although ultrahigh mass resolution is thus not essential, native MS still requires reasonably high instrumental resolving power to detect the individual charge states and mass shifts induced by PTMs on macromolecules or coexisting stoichiometries. Taken together, the results indicate that the best attainable resolution in native MS is analyte-specific and often pragmatically determined. Therefore, "High-Resolution" in the title of this review reflects the best

apparent (empirical) resolution attainable for the specific analyte rather than the inherently achievable resolving power of the instrument.

2.3. Mass Accuracy in the Detection of Macromolecules

Another crucial metric of performance, which is highly dependent on instrumental resolution, is the attainable mass accuracy. Mass accuracy is typically defined as the mass error as a fraction of the theoretical mass and is often expressed in parts per million or parts per billion (ppb)[54]. For small molecules (MW < 500 Da), depending on the attainable mass accuracy, a mass spectrometer can either unambiguously identify the molecule or point at a multitude of putative structures of similar yet distinct elemental compositions[55]. At the dawn of mass spectrometry, with mass analyzers having limited resolution and mass accuracy such as the quadrupole mass analyzer[56], it was only feasible to determine nominal masses, which are calculated by rounding masses of the most abundant isotopes[54]. However, with the advent of more advanced mass analyzers, such as ToF and, more notably, FT-ICR, it became feasible to measure masses incredibly close to the theoretical values, with sub-ppm or even ppb accuracy, albeit mostly at the cost of long observation periods[57]. While for the analysis of small molecules or for peptide/protein sequencing it is essential to determine monoisotopic masses, either directly or by inferring them from fine isotope distributions, in native MS average mass profiles are recorded. Not only is it challenging to obtain isotopically resolved spectra of pure macromolecular species (see **section 2.2**) but also the increasing probabilities of heavier isotopes in mass profiles of macromolecules result in the near absence of monoisotopic mass peaks[58]. As typical in native MS, recording average masses leads to more pronounced deviations from theoretical values, although these are typically still in the range of 2-30 ppm.

From a technical perspective, high mass accuracy requires a good mass calibration and elimination of all analyzer-specific systematic errors (discussed below). Calibration can be performed either internally or externally, with internal calibration, whereby the internal standard is present in every spectrum, yielding the best mass accuracies[45]. Calibrating the mass analyzer for native MS internally is challenging, and therefore, external calibration is more common. For that, most typically, a high-concentration aqueous CsI solution (about 10-100 mg/ml) is used, whereby the generated singly charged virtually monoisotopic CsI clusters can cover quite a wide *m/z* range and require very similar experimental conditions for gas-phase transmission and detection as large protein assemblies[59]. As for systematic errors of mass analyzers, they vary significantly since different mass analyzers adhere to distinct principles of mass analysis and detection. For example, becoming very popular in native MS, Orbitrap mass analyzers can suffer from space-charge effects[60] or electric field imperfections[61]. When all such systematic errors are eliminated, the mass accuracy is equal to mass measurement precision, reflecting random errors in multiple data points acquired for the same species[54].

The resolving power of the instrument usually defines the attainable mass accuracy. However, the empirical widths of peaks detected in native MS tend to often exceed the theoretical widths defined by the instrumental resolution[32]. Although careful sample preparation and dedicated ion cooling[62] in the mass analyzer can help reducing peak

widths, it is increasingly challenging to achieve this for large macromolecules. Since instrumental resolution is not the only limiting factor for accurate mass analysis of large ions with native MS, developments in mass analyzers have focused not only on aspects of pure mass detection but also on efficient manipulation of large macromolecular ions in the gas phase, maximizing analyte desolvation and adduct removal.

3. Features of Mass Analyzers Optimized for High-Mass Measurements

Although native MS has benefited massively from the general advances made in mass spectrometry over the past decades, the mass analyzers used for native MS still require distinctive characteristics, primarily because of the high molecular weights of the analyzed particles, their relatively lower charges, and the distinctive dynamics of ion motion within the mass analyzers. Initially, mass analyzers used for native MS were standard instruments designed for small molecules that were modified to enable the analysis of high-mass particles[63]. At the advent of native MS, modified ToF instruments were predominantly used, but more recently FT-ICR and Orbitrap mass analyzers have been adapted for high-resolution native MS[13,64].

Many modern implementations of these instruments feature a lower-resolution secondary mass analyzer, *e.g.*, a linear ion trap or a quadrupole, which is primarily used for ion selection, fragmentation, or other adjacent functions (**Figure 3**). Although essential for modern mass spectrometry, such secondary mass analyzers are often built-in along the transmission path of a mass spectrometer and may limit the attainable mass range, as their resolution and ion transmission efficiencies at high m/z values are not ideal. Nevertheless, modern implementations of quadrupoles with lower RF frequencies[65] or frequency-scanning (as opposed to amplitude-scanning) design[66], enable the extension of the attainable m/z range up to or even above 100 kTh.

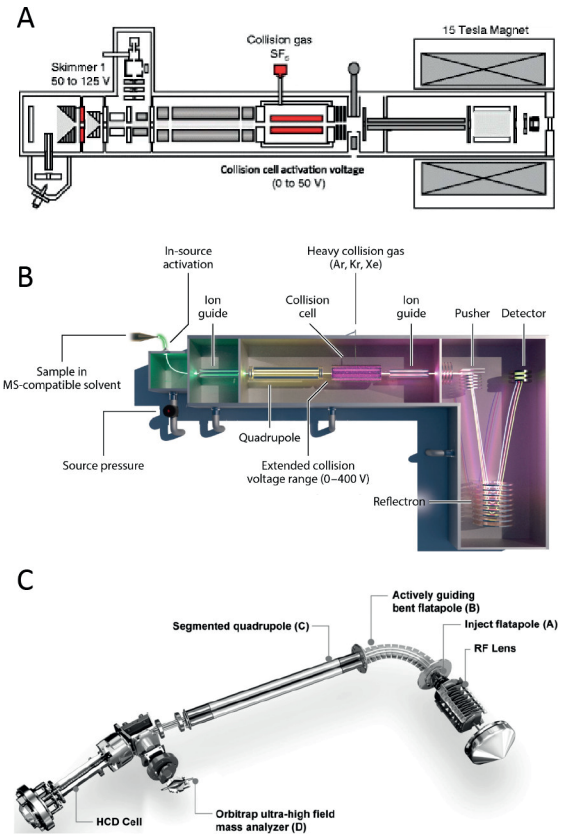


Figure 3 | Schematics of mass spectrometers commonly used for native MS. (A) 15 T FT-ICR. Adapted from ref [67]. Copyright 2017 American Society for Mass Spectrometry. **(B)** Synapt G2 instrument featuring a Q-ToF mass analyzer. Reproduced with permission from ref [68]. Copyright 2013 Nature Publishing Group. **(C)** Orbitrap Q-Exactive UHMR instrument with extended mass range. Reproduced with permission from ref [65]. Copyright 2017 Nature Publishing Group.

High-resolution mass spectrometers provide an array of useful and complementary sets of features that facilitate native MS analysis of large macromolecular assemblies. For example, ToF instruments enable fast scanning and a theoretically unlimited mass range[69], while FT-based mass spectrometers, although suffering from relatively slower signal acquisition, provide exceptional mass resolution, primarily in the low m/z range[38]. Novel Orbitrap-based mass spectrometers have been specifically tailored for sensitive measurements in the high-m/z region and provide increased practical resolution through superior desolvation[27,30,31,70]. Recently, with native MS increasing in popularity, a number of new ToF and FT-ICR instruments have emerged, improving upon initial limitations of speed and transmission in high-mass measurements[71-74]. Advances in attainable sensitivity, accuracy, resolution, and mass range over the past decade have empowered high-resolution native MS to unravel heterogeneity and structural features of very complex systems like membrane proteins[75], ribosomal particles[52], and crystalline oligomers[76], to name just a few.

3.1. Time-of-Flight Mass Analyzers

The ToF mass analyzers have proven to be extremely powerful in the analysis of large macromolecular assemblies and have been used to set mass records in native MS still unmatched by other instruments today, for instance, providing a charge-resolved mass spectrum of an intact 18 MDa bacteriophage assembly[34]. At its core, ToF mass analyzers allow for the simultaneous detection of masses in a very wide mass range, and furthermore, ToF-based analysis is inherently sensitive and fast[69]. Because in ToF mass analysis the m/z is proportional to the square of the time of flight of the ions through the analyzer, the resolution ($m/\Delta m$) of the ToF mass analyzer is equivalent to $t/2\Delta t$, which means that the resolution remains nearly constant over the whole m/z range (**Figure 4**). This feature is very distinctive from FT-based analyzers, where the resolution decreases drastically with increasing m/z.

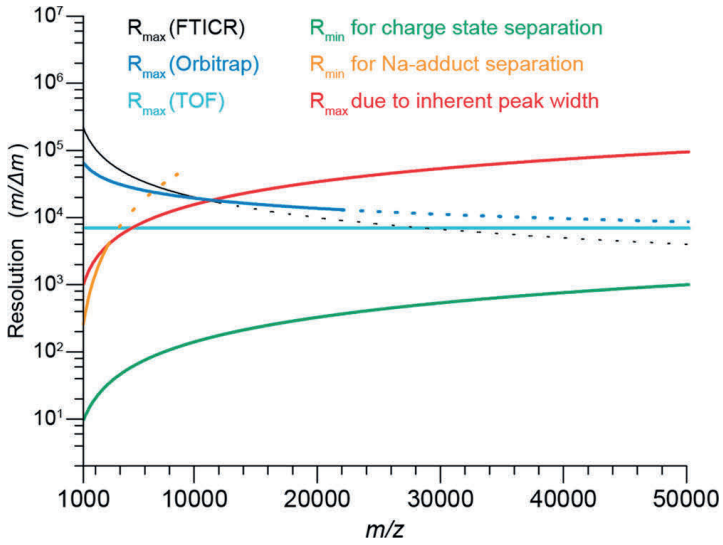


Figure 4 | Theoretical and experimental mass resolutions ($m/\Delta m$) in the m/z range from 1000 to 50,000 Th for the common mass analyzers used in native MS. Shown are plots of R vs m/z for the three mass analyzers (FT-ICR, Orbitrap, and ToF) based on their respective R_{max} values reported at m/z = 5000-6000 Th (FTICR, 40,000; Orbitrap; 25,000; ToF, 7000) and the theoretically achievable relations between analyzer mass resolution and m/z (FT-ICR (black line), $R \sim z/m$; Orbitrap (blue line), $R \sim (z/m)^{0.5}$; ToF (cyan line), R = constant). Dotted lines in the FT-ICR and Orbitrap graphs represent theoretical values for m/z ranges that were not accessible with these mass analyzers at the time of the publication. Reproduced from ref [32]. Copyright 2014 American Society for Mass Spectrometry.

Although the history of ToF analyzers goes back to 1948, when they appeared under the name 'Velocitron'[77], major developments important for native MS came much later. In the 1970s Mamyrin *et al*. introduced the ToF reflectron, which increased the resolution of ToF instruments by effectively extending the flight path and compensating for ion energy drift[78]. In the early 1990s, the advent of orthogonal-acceleration ToF (oa-ToF)[79,80] following innovations in fast digitizing electronics further improved the resolution of ToF MS by accumulating ions in the acceleration region and thus unifying their position and

velocity prior to acceleration. Shortly after the introduction of oa-ToF, the addition of an ion-selecting quadrupole (Q) compartment[81,82] enabled the conversion of ToF analyzers into so-called Q-ToF mass analyzers. The first Q-ToF mass spectrometers used for native MS featured nano-ESI (nESI) sources and were modified to enable an increased pressure regime in the instrument's front end. While nESI was essential for soft ionization of macromolecules, high pressure was crucial for ion cooling through collisions with neutral gas molecules, which is absolutely necessary for electrostatic focusing of large ions[62]. Various solutions for efficient ion cooling were explored by both the Heck and Robinson groups between 2006 and 2009, including restriction of pumping in the front end of the mass spectrometer[83,84] and the use of heavy gases in the collision cell[85]. For instance, several important modifications to employ Q-ToFs for the analysis of macromolecular assemblies were implemented by Van den Heuvel *et al*. in 2006[86]. They modified a first-generation Q-ToF by introducing high-transmission grid ion optics, a low-frequency quadrupole, a low-repetition pusher, and a high-pressure collision cell. These developments enabled the pioneering mass measurements of large viral particles[87-89] and membrane protein assemblies[90,91], exposing – in addition to the MW – their structural features, such as composition, protein-ligand stoichiometry, and binding stability.

In the past decade, developments in ToF instruments have primarily focused on tailoring the electronics for ion-selecting or ion-routing optics and further gas regime control to improve the transmission of high-mass ions. Since high resolution requires increasing the flight distance of ions between the pusher and detector, multiple solutions have recently emerged to improve the mass resolution, including multipass and multireflection instrumental setups[92-96]. These recent experimental improvements have led to significant improvements in resolution (>10,000) and mass accuracy (~ 5-10 ppm) while simultaneously improving the transmission of high-mass ions and spectral quality in ToF-based native MS. Still, on such instruments the ion signals recorded by native MS remained much broader than would be expected if they were purely limited by the mass resolution of the instrument, resulting often in an overestimation of molecular weights by up to a few percent[32]. As discussed above, this limitation originates primarily from imperfect desolvation of the ions in transmission-type Q-ToF instruments. In addition, the extended ion flight paths in multireflection and multipass instruments additionally imply lower sensitivity in detecting macromolecular assemblies, which may dissociate prior to reaching the detector. If such hurdles can be addressed, ToF-based mass analyzers would further consolidate and could possibly see increased usage in native MS.

3.2. FT-ICR Mass Analyzers

Since its introduction back in 1974[97], FT-ICR MS has developed into a wonderful tool for mass measurements enabling ultra-high mass resolution (> 10^6 @ m/z = 200 Th)[98] and ultra-high mass accuracy, even reaching into the ppb range[57], significantly advancing the fields of proteomics, metabolomics, petroleomics and environmental analysis, among others. Nevertheless, for a long time, the application of FT-ICR for native MS was somewhat limited and only started to be substantially explored around a decade ago with the emergence of instruments using strong field magnets (*e.g.*, 12 or even 15 T)[99].

FT-ICR instruments represent a variation of a Penning trap device, which traps ions along the axis of the ICR cell within a spatially uniform magnetic field (B) that provides radial ion trapping and governs their cyclotron motion[38]. This ion cyclotron motion in turn induces an image current in the two opposing electrodes when the ions are excited to higher orbits. Signal detection in FT-ICR can happen only after the ion excitation step, enabling ions with identical m/z to rotate coherently with a locked phase close to the detector electrodes. Facilitating mass detection, ion cyclotron resonance is independent of the ion energy, and the recorded signal is linearly proportional to the rotational amplitude of ion motion. Such linearity enables the time-domain signal to be recorded and subsequently converted it into a frequency-domain spectrum by Fourier transformation[100], adapting FT principles developed within the rich field of Fourier spectroscopy and NMR spectroscopy[101]. Since the ion "image current" is induced by ions rotating at a cyclotron frequency $f_c = B/[2\pi(m/z)]$, resolution in FT-ICR is directly dependent on the number of cyclotron rotations and, therefore, the length of the ion flight path. This holds true only when ions of the same m/z are coherently locked in phase within a perfectly homogeneous magnetic field.

The advantages of FT-ICR for high-resolution measurements compared to ToF become clear when their corresponding lengths of ion flight paths are compared. For example, ions with m/z = 1000 Th analyzed by FT-ICR with a 9.4 T magnet in 1 second undergo 144 346 rotations, which is equivalent to a flight path of 9 km[102], an incredible leap compared with the 2-5 m ion flight path in a typical ToF instrument. However, the outstanding resolving power comes at a high cost. Since the cyclotron frequency is proportional to $(m/z)^{-1}$, the attainable instrumental resolution drops rapidly with increasing m/z, which somewhat limits the usability of FT-ICR in native MS of larger protein assemblies. The direct dependence of the resolution on the duration of the time-domain signal also hampers high-resolution detection of large macromolecules. Their ions signals are often susceptible to decay due to magnetic field imperfections, collisions with neutral background gas molecules, and coherence-disrupting ion-ion interactions.

Since ion spatial coherence is of key importance in FT-ICR MS, it was crucial to develop methods for pulsed rather than continuous introduction of ions into the ICR cell. Therefore, in the 1990s, FT-ICR instruments were equipped with a multipole, which enabled ions to be accumulated and then simultaneously injected as a packet into the ICR cell[103]. This modification enabled the efficient excitation of ions and phase locking. Another substantial improvement of ICR-type instruments was achieved by increasing the magnet conductivity. It was shown that with stronger magnetic fields multiple metrics of FT-ICR performance could be enhanced, with some (*e.g.*, the upper mass range) scaling quadratically[104]. Modern FT-ICR instruments can be equipped with very powerful 15 T or even > 20 T magnets. These higher magnetic field strengths not only increase the attainable resolution, accuracy of mass measurement[57], and stability of ion cyclotron motion in the ICR cell, but also initiated the application of FT-ICR for the analysis of large biomolecules, especially for the analysis of intact antibodies[99,105,106]. Apart from using more conductive magnets, multiple ICR cell designs have been proposed to reduce magnetic and electric field imperfections and increase the mass resolving power. Some of the recent advances were reported by Campuzano *et al.*[107], who described high-resolution

native MS analyses of antibody-drug conjugates (ADCs) using a powerful 15 T solariX FT-ICR instrument (**Figure 5**).

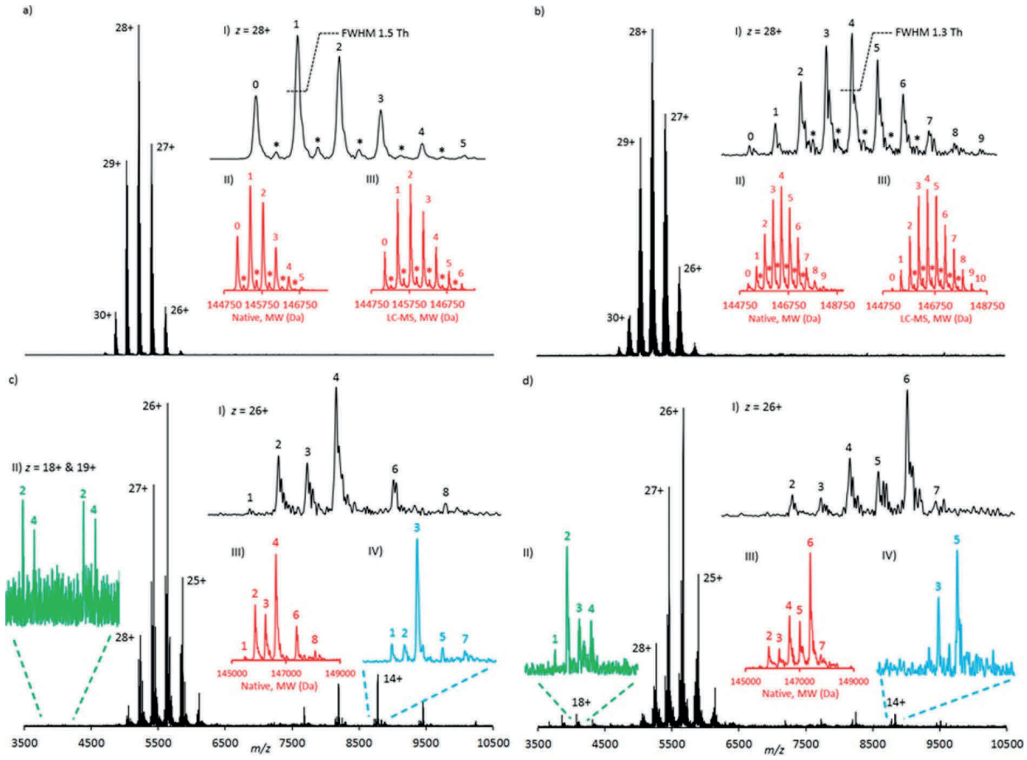


Figure 5 | High-resolution native MS analyses of antibody-drug conjugates using a 15 T FT-ICR instrument. Shown are the native MS spectra of **(a)** 5 equiv (biotin) lysine conjugate, **(b)** 10 equiv (biotin) lysine conjugate, **(c)** 5 equiv tris(2-carboxyethyl)phosphine (TCEP) and 10 equiv (biotin) cysteine conjugate, and **(d)** 10 equiv TCEP and 10 equiv (biotin) cysteine conjugate. Drug-to-antibody ratio (DAR) values are annotated on selected charge states and all of the deconvoluted spectra. Asterisks (*) represent a +162 Da glycation mass increase. Data are displayed in magnitude mode using a symmetric Hann apodization function. The annotated DAR values in insets III and IV of (c) and (d) are consistent with the numbers of biotin moieties covalently attached to the specific species and not the full intact initial monoclonal antibody conjugate molecule. Peak widths at half-height for z = 28+ are annotated. The peak widths at half-height for the equivalent deconvoluted data range from 39.2 to 41.1 Th. Reproduced from ref [107]. Copyright 2017 American Chemical Society.

The high-resolution capabilities of modern ICR mass spectrometers make them well-suited for native top-down MS analysis of proteins and their assemblies, whereby the isotopic envelopes of low-charge dissociation products can still be resolved. Because of the required ultrahigh-vacuum operation, FT-ICR instruments are less compatible with conventional collision-based ion fragmentation methods, such as collisional dissociation (CID). Therefore, alternative fragmentation methods have been introduced, featuring – among others – photon-induced and electron-induced fragmentation techniques, which also were shown to have benefits over CID in the structural investigation of larger protein assemblies[28,67].

Overall, high-field FT-ICR instruments represent excellent mass analyzers for native MS, although their advantages come at the expense of a relatively high cost, which – together with their relatively large size and high maintenance – likely has hampered somewhat the wide usage of these instruments in the field to date.

3.3. Orbitrap Mass Analyzers

Despite being the youngest among the mass analyzers frequently used for native MS, the Orbitrap has experienced rapid developments and currently represents one of the most prominent technologies for the mass analysis of large and heterogeneous macromolecular assemblies[30,70,75,108]. The Orbitrap typically forms the end-stage detector of a hybrid mass spectrometer that is further equipped with other mass analyzers such as quadrupoles and ion traps. Through the combination of these mass analyzers, Orbitrap-based instruments can build upon several strengths of more mature mass analyzing/manipulating devices while eliminating some of their drawbacks[109]. Ultimately, Orbitraps inherited many features, such as pulsed ion injection from the earlier developed oa-ToF, trapping principles from radiofrequency ion traps, and signal detection and processing from FT-ICR. Circumventing some of the shortcomings, the Orbitrap mass analyzer does not require superconductive magnets and benefits from significantly smaller dimensions compared with ICR cells while enabling "ion excitation by injection"[61]. In addition, Orbitraps experience a less steep drop in resolution with increasing m/z, as the oscillation frequency – in contrast with the cyclotron frequency in FT-ICR – scales proportionally to $(m/z)^{-1/2}$ (**Figure 4**). Similar to FT-ICR and as an improvement upon the ToF mass analyzers, the recorded axial ion oscillation frequency in an Orbitrap is independent of the ion energy, making the Orbitrap very suitable for mass detection.

Physically, the Orbitrap can be represented as a two-electrode device consisting of an inner spindle-like electrode and an outer curved electrode, which together create a quadrologarithmic field[110]. Although the Orbitrap was introduced in 1999 and commercially released as a linear ion trap Orbitrap (LTQ Orbitrap) instrument in 2005, the principles of orbital ion trapping were established by Kingdon about a century ago[111]. In an Orbitrap, ions oscillate around the central electrode with m/z-specific axial frequencies, inducing a current on the split-in-half outer electrode. Prior to mass detection, ions are first stored and confined in a linear ion trap adjacent to the Orbitrap, known as the C-trap, and then injected into the mass analyzer by applying a voltage onto the back plate of the C-trap[41]. In contrast to FT-ICR, such external excitation is sufficient to achieve spatial coherence of the ions without additional oscillating or rotating electric fields, although dipolar excitation has been explored for selective ion ejection and motion stabilization[112,113].

Since its introduction, the Orbitrap mass analyzer has undergone several developments, including the introduction of a high-field Orbitrap mass analyzer and enhanced methods for processing of the time-domain waveform signal[114], but arguably the most important modifications for native MS were those targeting the ion optics and electronics. High-field Orbitrap mass analyzers feature a more compact design with improved ion trapping and enhanced resolution and dynamic range[115]. The utility of the Orbitrap for native MS was

first demonstrated in 2012 by Rose *et al*., who modified a benchtop Orbitrap instrument by lowering the radiofrequencies of the ion transition optics and implemented a gas control system for the higher-energy collisional dissociation (HCD) cell[27]. With these modifications, the Orbitrap instrument demonstrated efficient transmission and collisional cooling of large (high m/z) ions, outperforming Q-ToF instruments available at that time in terms of resolving power but not yet in mass range. The high resolution and sensitivity of the Orbitrap mass analyzer provided a means for detailed analysis of viral particles[116] and biopharmaceuticals with native MS[31,117]. A schematic of a modern Orbitrap-based mass spectrometer for native MS is provided in **Figure 3C**.

Shortly after this first demonstration, commercial Orbitrap mass spectrometers were introduced, geared toward analysis in the extended mass range (Orbitrap-EMR). This first generation of instruments lacked possibilities for mass selection, which is essential for tandem MS analysis of large analytes in native MS. This problem was overcome by Belov *et al*., who equipped the instrument with an EMR quadrupole operating at lower radiofrequencies[118] to enable selection of high-mass ions. Dyachenko *et al*. shortly thereafter demonstrated the utility of these innovations by measuring stoichiometries and localizations of drug moieties in ADCs and antibody-antigen complexes exhibiting masses in the MDa range[119]. More recently, an Orbitrap with ultrahigh mass range (UHMR) capabilities that was developed for native MS was commercially released, closing the gap in upper mass limit with ToF MS and providing even more powerful desolvation and ion cooling capabilities in the front end of the instrument[30,65]. The Heck group has used such instruments to stretch the boundaries of accurate high-resolution mass analysis up to m/z = 50,000 Th by studying intact Flock House viruses[65]. The ultra-high mass range allowed van de Waterbeemd *et al*. to study compositional variants and to unravel heterogeneity within ribosomal particles extracted from three different organisms[52]. Multiple groups also utilized the strong desolvating capabilities of the instrument to analyze membrane proteins[21,120] and even to perform triple-stage tandem MS analysis, wherein fragmentation in the front end is followed by product isolation using the quadrupole and secondary fragmentation either by conventional HCD[70,75] or custom fragmentation methods (*e.g.*, surface-induced dissociation by the Wysocki lab[121] or UV photodissociation by the Brodbelt[122] and Heck groups[123]). Among the latest advances, Orbitraps were utilized for ion mobility analysis, providing an extra dimension to reduce sample complexity and further advance structural analysis of macromolecules[124,125]. These recent developments demonstrate that the future may still bring many more exciting possibilities for Orbitrap-based native MS, some of which will be discussed further below.

4. Novel Solutions for Enhancing Resolution and Disentangling Heterogeneous Macromolecules by Native MS

The demand for mass analyses of heterogeneous macromolecules with ever-increasing masses is continually expanding, bringing additional challenges to native MS. Since, as described in the earlier sections, the instrumental mass resolution is not the major practical limiting factor for accurate mass detection in native MS, most of the recent advances in resolving high-mass and high-heterogeneity samples are focused on other factors such as improving gas-phase transmission and desolvation (see **section 3**), advancing algorithms for data analysis, and implementing separation in dimensions orthogonal to m/z. Additionally, newer approaches such as single-ion mass spectrometry can boost empirical resolution, circumventing high sample microheterogeneity issues, especially for high-mass analytes[126], by the use of charge-detection mass spectrometry (CDMS)[127]. In this section, we further review these developments in more detail.

4.1. Advanced Algorithms for FT MS

In FT MS, high resolution comes at the expense of the need for long ion observation times. This is exacerbated by imperfect magnetic and electric fields in FT-based mass analyzers, which result in gradual dephasing or loss of coherence in the trajectories of the ions. Moreover, while ion cooling is beneficial, this is typically performed by increasing the pressure along the transmission path, but the presence of gas molecules is detrimental in the acquisition of long transients. To tackle such limitations, it is desirable to reduce the ion observation time without decreasing the resolving power. In 2013, for Orbitrap MS, this was partly achieved by improving upon the initial data processing with the advent of enhanced Fourier Transform (eFT), with benefits of both absorption and magnitude-mode signal processing[114], whereby absorption mode FT provides a resolution nearly twice as high as that of magnitude mode, which in turn results in more robust amplitude spectra. Absorption-mode processing was also implemented for FT-ICR instruments, although relatively late, as in FT-ICR it is challenging to achieve phase coherence for all of the analyzed ions[40,128]. Even more recently, a new method for FT MS data processing was introduced, named ΦSDM, which is based on deconvolution of FT spectra beyond the FT uncertainty, enabling high resolution by the use of just a small fraction of transient length when compared to eFT[129]. Although it has been implemented only on Orbitrap mass analyzers and optimized primarily for the low m/z region, for instance as used in TMT labeling experiments[130], ΦSDM has the potential to improve the detection of short-lived signals characteristic of ensembles of also large macromolecules as measured in native MS, thus providing improved resolution even when short transients are recorded. This, however, still needs to be demonstrated in practice.

4.2. Additional Dimensions for Gas-Phase Separation in Native MS

An alternative way to reduce sample complexity and increase resolving power is by separating ions on the basis of their mobility in the gas phase. We will not discuss ion mobility MS (IM-MS) here extensively since it is covered in depth by another review in this issue and has already been reviewed in earlier publications[131-133]. However, it is worth mentioning that IM-MS is a powerful and rapidly evolving branch of mass spectrometry in general, with fruitful features also for native MS. It not only provides a means to separate complex mixtures of proteins but can also be used to probe structural features. To illustrate the latter, IM-MS was recently used in a wide range of structural applications, including investigations of amyloid formation[134], conformational landscapes of the ubiquitin protein[135], and structural features of DNA/RNA telomeric G-quadruplexes[136]. Moreover, combining ion mobility with tandem mass spectrometry has proven to be useful for fingerprinting of biopharmaceuticals[137], and studies of protein interactions[138], and conformational dynamics of protein assemblies[139].

Spatial separation of ions was recently explored by Mathew *et al*.[140], who coupled a modified ToF instrument to a Timepix pixelated detector, which enabled them to unravel distinctive trajectories of ions in the mass spectrometer upon excitation. Such a hybrid instrument provided a first example of single-ion imaging of protein assemblies in native MS and distinguished between the signal of protein ions and secondary electrons produced by ion-surface collisions. The time- and space-sensitive detection with the Timepix detector has the potential to facilitate the disentangling of dissociation products of protein assemblies in native top-down MS.

4.3. Single-Molecule and Charge Detection for Analysis of Large and Heterogeneous Assemblies

Aside from tackling instrumental and processing resolution limits, alternative experimental designs have also been explored to improve the performance of native MS, most notably by measuring single ions as opposed to ensemble ion detection. Single-ion signals can be statistically postprocessed, enabling the data to be filtered to remove aberrant ion signals, thereby improving the S/N, mass resolution, and accuracy. This is possible because each ion is analyzed separately, and its detection is thus void of space-charge effects, ion coalescence, or high signal damping, which makes it more similar to time-resolved ion detection using time-to-digital converters in ToF MS. However, compared with ToF MS, though, for which single-molecule detection in native MS was already demonstrated in 2000[141], FT MS approaches provide much higher resolution and mass accuracy.

Single-molecule analysis in native MS can be performed using an Orbitrap mass analyzer. Single-molecule detection was initially demonstrated for intact myoglobin[142] and readily showed superior resolution and mass accuracy compared with conventional ensemble measurement. Shortly after, a method utilizing multiple single-molecule scans to enhance resolution and accuracy was proposed by Makarov[143]. As a proof-of-principle for measuring single ions in native MS, Rose *et al*.[27] demonstrated single-molecule detection

for the intact 800 kDa GroEL protein complex on the Orbitrap EMR. More recently, using single-molecule analysis and postacquisition spectral filtering to remove decaying ion signals, Kafader *et al*.[126] achieved a 10 to 20-fold resolution improvement for intact proteins with masses of up to 150 kDa on an Orbitrap instrument. Overall, single-molecule analysis extends far beyond conventional native MS and has been implemented using different ion detection principles, which are further described elsewhere[127].

In addition to improved mass resolution and accuracy, single-molecule analysis allows researchers to probe the charge states of proteins that are typically too large or too heterogeneous to analyze with the conventional native MS approach. Early on, CDMS experiments by the Smith group[144,145] on FT-ICR instruments enabled the detection of extremely large ions (> 100 MDa) with a huge number of carried charges (> 30,000), such as T4 DNA particles[146]. Since then, CDMS has become a useful tool that is complementary to native MS for measuring large ions and has been implemented on specialized platforms, often featuring an electrostatic linear ion trap (ELIT) with a detecting cylinder inside[147]. For a long time, resolution in CDMS devices was lagging. However, recent advances have shown significant improvements, as exemplified by Jarrold's group, who combined dynamic calibration and advanced ELIT designs[148]. The lower limits typically restrict CDMS to 250 charges and 1 MDa for mass. However, most recently, several exciting CDMS applications in the mass range similar to that in conventional native MS have been achieved by using Orbitrap UHMR mass analyzers, wherein single-ion signal amplitudes are proportional to the charge. Such data were reported simultaneously by the groups of Heck[108] and Kelleher[149], whereby the new charge dimension allowed them to disentangle extremely heterogeneous viral particles and large immunoglobulin oligomers with high charge and m/z accuracy.

5. Spectral Deconvolution in High-Resolution Native MS

Native MS spectra can be very dense and congested because of both heterogeneity of the samples and often co-occurring ion signals of multiple charge states of the same species. In isotopically resolved spectra, overlapping ion signals of different ions and charging by charge carriers other than protons may further complicate the spectra. Therefore, the development of spectral deconvolution tools has become crucial for the interpretation of such native MS data. In this section we review these developments, primarily focusing on deconvolution in native MS but also addressing the deconvolution of isotopically resolved mass spectra.

5.1. Spectral Deconvolution for Native Mass Spectra

In ESI-based MS, biomolecules are detected as gaseous ions with a distribution of charge states, represented by adjacent peaks in the m/z space. Provided that the mass resolution is sufficient to record charge-resolved spectra, mass determination requires that these distributions of multiply charged ion peaks are converted into a monodispersed zero-charge format. Over the past three decades, multiple algorithms have been developed to

address this either through specific peak assignment in the charge distributions or by simulation and fitting of a hypothetical spectrum to the entire raw spectrum or parts thereof. While the former category provides speed and robustness for the deconvolution of relatively simple spectra, the latter excels at addressing challenges in the mass analysis of heterogeneous samples, although these algorithms are generally more computationally intense.

The first algorithm for automatic charge state assignments and mass deconvolution of multiply charged mass spectra was developed by Mann *et al*.[150] at the advent of electrospray ionization more than 30 years ago. In essence, the algorithm achieved mass determination by iteratively calculating the cumulative charge state abundance as a function of a hypothetical mass in a defined range. Although it was efficient for deconvolution of uncomplicated spectra of pure proteins, this algorithm suffered from some artifacts and was less useful for the deconvolution of spectra of mixtures. The next generation of deconvolution algorithms appeared in the early 1990s and utilized a metric to ensure the best probability-based fit between the predicted and experimental charge distributions[151,152]. Maximum-entropy-based algorithms, in particular, resulted in a more reliable deconvolution. However, some drawbacks persisted, especially regarding speed, quantitation, spectral complexity, and deconvolution artifacts. Over the next two decades, various implementations of the algorithms, which combined advanced peak picking and charge assignment, culminated in overall faster and more accurate deconvolution algorithms[153-155]. In parallel, novel approaches based on finding the best fit to raw data among multiple simulations of mass and charge distributions were pushing the boundaries of disentangling spectral crowdedness[156,157], as exemplified by the robust tool Massign developed by the Robinson group[158]. The difficulty of deconvolving overlapping charge distributions, which is common for multimeric heterogeneous protein assemblies, was further optimized by using the minimax theorem originating from game theory (AutoMass)[159] and – to an extent – resolved by second-derivative-based peak detection (PeakSeeker)[160]. Although direct peak detection provided significant advantages for disentanglement of complex native mass spectra, it often failed to use all of the peaks in the experimental mass spectra, which became a large focus of the more recent emerging tools for spectral deconvolution in high-resolution native MS.

In 2015, Marty *et al*. released UniDec (short for Universal Deconvolution), a software tool based on a Bayesian deconvolution algorithm that enables fast and quantitative processing of native MS data[161]. Although the approach relied on some user input for correct mass deconvolution of complex mass spectra, it provided significant improvements in speed and accuracy compared with earlier methods, chiefly by applying a customized Richardson-Lucy deconvolution algorithm with separate charge and mass smoothing. In addition, UniDec allowed for the processing of native ion mobility mass spectra along with native MS data of varying complexity. Unlike many other academic deconvolution tools, UniDec has been well-maintained and further developed over the years. It is worth mentioning a few later additions, among which are an extension for batch processing of large data sets (MetaUniDec)[162] and the SoftMax function for removal of artifacts in deconvolution of extraordinarily congested and charge-overlapping spectra[163] (**Figure 6**). In 2020, Marty presented a universal scoring system for quality assessment of

each deconvoluted mass feature[164], further improving the spectral deconvolution reliability in native MS.

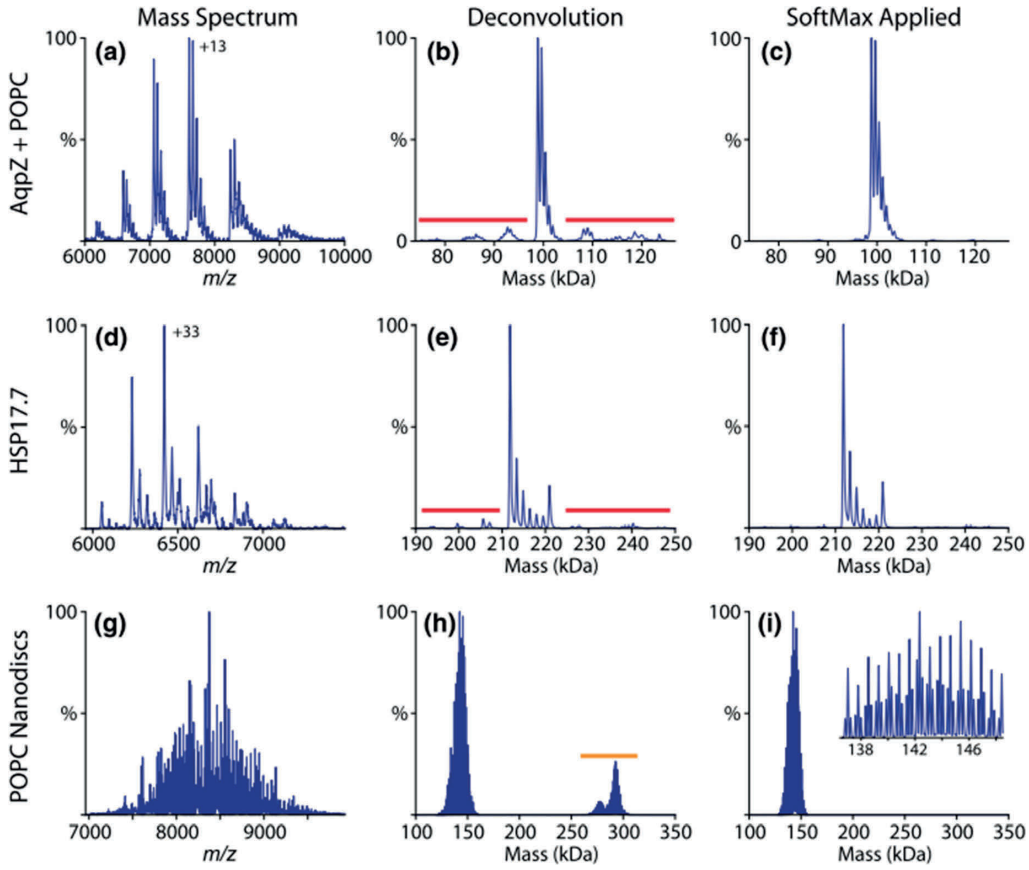


Figure 6 | Spectral deconvolution of native MS spectra using UniDec software. The native mass spectra of AqpZ with bound POPC **(a)** and HSP17.7 **(d)** show satellite artifacts (red) when deconvolved **(b, e)** that are removed by addition of the SoftMax function **(c, f)**. POPC nanodiscs with mixed heavy and light belts **(g)** show harmonic artifacts (orange) after deconvolution **(h)** that are removed by the SoftMax function **(i)**. The triplet peaks from the mixed belts are preserved by the SoftMax function as shown in the inset. Reproduced from ref [164]. Copyright 2019 American Society for Mass Spectrometry.

Commercial software solutions for deconvolution seem to be somewhat biased, being primarily geared for the biopharmaceutical industry[165], showcasing robust performance for the analysis and quality control of biotherapeutic molecules such as antibodies and ADCs. Two of the currently most prominent algorithms include ReSpect, which uses an implementation of the maximum-entropy algorithm and is available in the BioPharma Finder software (Thermo Scientific), and another commercial deconvolution tool called PMI Intact (Protein Metrics)[166]. Within the BioPharma Finder suite, ReSpect is primarily used for the deconvolution of intact protein LC-MS data. On the other hand, in a recent study by Campuzano *et al.*, PMI Intact has been shown to cope exceptionally well with very

complex high-resolution native mass spectra of common biopharmaceuticals and polydisperse nanodiscs[165] (**Figure 7**). In addition to advanced peak picking and smoothing of deconvoluted data, the PMI Intact algorithm uses a "parsimonious" criterion, which ensures that the resultant zero-charged spectrum has the minimum number of peaks necessary to explain the underlying experimental MS data.

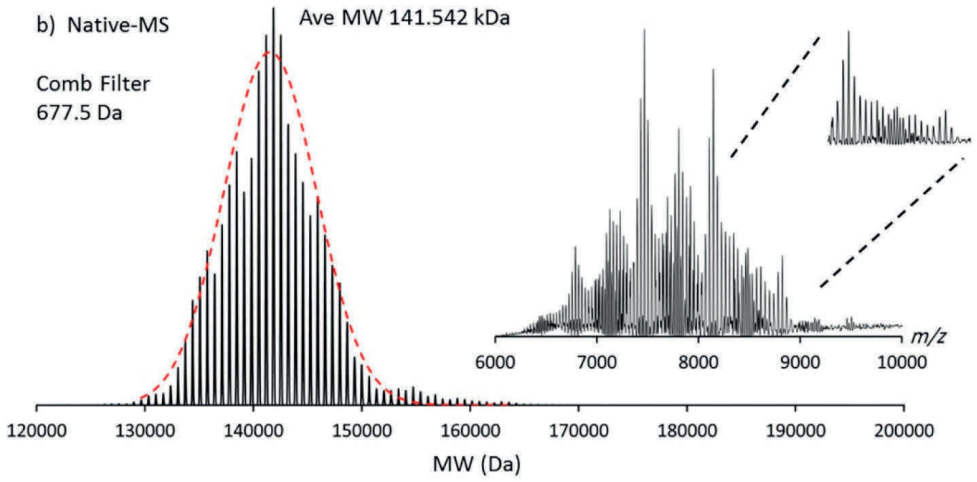


Figure 7 | Spectral deconvolution by PMI Intact software. Shown is the deconvolution of the highly polydisperse native mass spectrum of an empty MSP1D1 nanodisc containing the phospholipid DMPC analyzed on an Orbitrap EMR. The inset displays the initial native mass spectrum. The Comb filter, available in PMI Intact, results in the successful deconvolution of this complex spectrum. Reproduced from ref [165]. Copyright 2019 American Chemical Society.

Two alternative recent approaches for the deconvolution of complicated mass spectra have emerged from the Prell and Kohlbacher laboratories. Using a fast Fourier transform-based algorithm allowed the Prell lab to deconvolve heavily populated spectra of heterogeneous ion populations while practically eliminating the need for prior knowledge about the putative mass and charge ranges and subunit composition. Potentially, this promises to be a more parameter-free method compared with existing deconvolution tools[167]. By showcasing the algorithm's strength on complex spectra of "empty" nanodiscs (**Figure 8**), Cleary *et al*. demonstrated how overtone peaks in Fourier spectra could complement Bayesian deconvolution as implemented in UniDec[168]. It is worth noting that successful analysis with this approach requires repeating mass units to be present in the spectrum. Another novel deconvolution algorithm was recently published by the Kohlbacher lab, whereby they achieved speedy and robust charge assignment by transforming the observed raw peaks into the logarithm of m/z, taking advantage of the resultant charge-specific and mass-independent constant patterns that could be easily identified[169]. It is exciting to see all these parallel developments in spectral deconvolution that, combined with advances in mass resolution and mass range, promise to improve mass identification in high-resolution native MS, potentially extending its applications to previously uncharted areas of research.

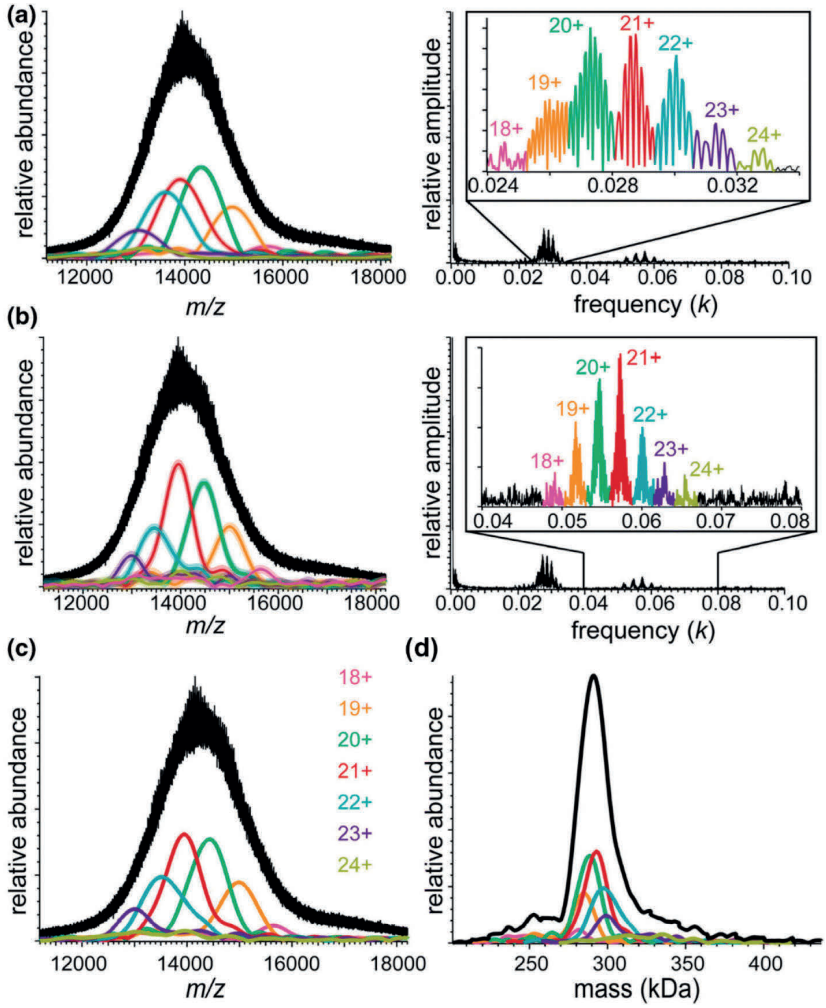


Figure 8 | Native mass spectra of DPPC-MSP1E3D1 nanodiscs acquired on a Q-ToF mass analyzer and corresponding Fourier spectra for **(a)** the fundamentals and **(b)** the second harmonics. Inverse Fourier transforms of the charge-state-specific peaks in Fourier spectra (insets) are shown as overlaid envelope functions of the same color in the mass spectra. **(c)** Harmonic-averaged reconstruction of envelope functions. **(d)** Zero-charge spectrum (black), calculated from harmonic-averaged spectra for all charge states. Reproduced from ref [168]. Copyright 2018 American Society for Mass Spectrometry.

5.2. Spectral Deconvolution for Isotopically Resolved Mass Spectra

Most of this review is focused on the analysis of large proteins and protein complexes by native MS, whereby the analytes cannot be isotopically resolved (yet). However, for sufficiently small proteins, typically below 50 kDa, isotopically resolved spectra can already be acquired using high-resolution mass analyzers, especially those based on FT MS and some of the more recently introduced high-resolution ToF-based instruments. A few groups are currently aiming to extend the reachable range for isotopically resolved spectra up to 150-200 kDa, either by single-ion measurements[126], or by long-transient

acquisitions as possible on high-end FT-ICR instruments[72]. Advantageously, in isotopically resolved spectra the charge can be directly extracted from the *m/z* differences between consecutive peaks in an isotope envelope. Moreover, this approach allows for either explicit or implicit detection of the monoisotopic masses, whereby implicit detection using robust computational approaches is essential for large molecules, for which monoisotopic mass peaks are nearly undetectable (*e.g.*, 0.04% relative abundance for ~ 17 kDa species)[58]. The application of spectral deconvolution methods in native (top-down) MS is exceptionally pertinent in protein fragmentation, whereby some of the backbone fragments are often detected in a single charge state, making high-resolution acquisition (> 50,000 at *m/z* = 200 Th) and, therefore, deconvolution of isotopically resolved spectra a necessity[170]. As native MS is more and more combined with tandem MS[28,171], we also include here a brief discussion about progress in spectral deconvolution of isotopically resolved mass spectra. Such deconvolution has been extensively developed in parallel with algorithms for unresolved mass spectra, with tools like Zscore[155], ProMex[172], and FLASHDeconv[169] supporting both types of data as an input. Specific algorithms have been explicitly tailored to handle highly charged and congested spectra of isotopically resolved protein fragments, including MS-Deconv[173], YADA[174], THRASH[175], and its commercial implementation Xtract available in the BioPharma Finder software suite (Thermo Scientific). More recently, the Liu group released EnvCNN, a statistical AI-based model for scoring identified isotopic envelopes, whereby they demonstrated superior performance compared to existing scoring functions[176]. While these tools are primarily used in conventional top-down MS analysis, most of them can also be applied to the deconvolution of native top-down mass spectra[28,171].

Overall, there are many tools available for spectral deconvolution of high-resolution native mass spectra, which have substantially helped the field make the advances described in this review. Notwithstanding this huge progress, quite a few challenges remain ahead. For instance, deconvolution of spectra on the fly (in real time), which has been achieved to an extent on some Orbitrap-based instruments, will significantly facilitate the optimization of data acquisition methods. Another challenge for the upcoming deconvolution methods is to handle incomplete desolvation and salt adducts, which are the most common sources of errors in mass determination in native MS. As not yet widely implemented, deconvolution can be improved by better leveraging of the expected repeating mass differences in native mass spectra, stemming, for instance, from the infamous salt adducts. Finally, with the advent of CDMS, there is a huge potential in utilizing additional layers of information about single ions to guide deconvolution algorithms for more accurate results.

6. Tandem Mass Spectrometry and Ion Activation in Native MS

Tandem mass spectrometry, or MS/MS, provides a means to obtain an additional layer of structural information about the analytes and is used widely to sequence peptides and small proteins. In such experiments, the precursor ions are mass-selected and submitted to a particular type of ion activation with the aim to fragment them. Tandem MS thus requires the deposition of energy into the precursor ions, and in the field of mass spectrometry there are currently quite a few different ion activation methods available[177-179]. Although these methods have not been developed with native MS applications in mind, they are often complementary, with some being more applicable to native MS than others, also depending on the aim of the ion activation. Often in native MS, ion activation is not necessarily used to generate dissociation products but rather is employed to improve desolvation[62], enhancing the attainable mass resolving power. This section briefly describes first the different ion activation methods available and subsequently their particular applications in native MS.

Ion activation techniques can be broadly categorized into collision-based methods, irradiation-based methods, and methods based on ion-ion interactions[180]. Collisions with inert gas molecules (collision-induced dissociation (CID) or collisional-activation dissociation (CAD)) represent the most frequently used ion activation methods in mass spectrometry, including native MS. This dominance likely originates from their widespread availability on nearly every mass spectrometer as well as their simplicity and robustness[181]. Although displaying excellent fragmentation efficiency, collision-based activation is mainly hampered by slow and stepwise energy deposition that primarily dissociates the most labile bonds, often leading to small neutral losses and other less informative dissociation products. Over the years, alternative ion activation techniques have been introduced that complement CID/CAD and provide better access to high-energy dissociation pathways. These alternative activation methods include methods based on collisions with surfaces (surface-induced dissociation (SID[121])), interactions with electrons (*e.g.*, electron-capture dissociation (ECD[182]) and electron-transfer dissociation (ETD[183]), and through the absorption of either low-energy photons (infrared multiphoton dissociation (IRMPD)) or high-energy photons (ultraviolet photodissociation (UVPD[184])). Because of the cross-complementarity of different ion activation methods, multiple hybrid techniques have recently emerged, with some getting widely adopted by the mass spectrometric community, *e.g.*, electron-transfer higher-energy collisional dissociation (EThcD), in which ETD is combined with supplemental low-energy collisional activation[185,186]. We refer readers interested in a more detailed overview of ion activation technologies in mass spectrometry to excellent reviews from the Brodbelt group[179,187].

When applied to protein complexes studied by native MS, ion activation methods are used to (1) enhance desolvation and improve mass resolving power, (2) eject protein subunits from non-covalent protein complexes to provide structural information, and (3) fragment polypeptide chains of individual subunits to determine amino acid sequences used to identify the protein. While the slow, stepwise, and low-energy activation provided by collisional activation works well for desolvating ions and removing adducts (1), faster and

more energetic photon- or electron-induced methods are more suitable for fragment formation and sequencing of subunits from within native complexes (3). Non-covalent dissociation (2) can benefit from various techniques, as distinct mechanisms of protein complex partitioning provide very complementary results.

6.1. Improving Mass Resolution and Accuracy by Ion Activation through Enhanced Desolvation

Incomplete removal of solvent adducts (*i.e.*, water, buffer, and salt molecules) is a major limiting factor for obtaining high resolution and mass accuracy in native MS, as this increases the peak width of the detected ion signals and artificially increases the measured mass[32,188]. Carefully balancing the amount of ion activation allows residually bound small molecules to be removed, whereby the internal energy should remain below the thresholds for fragmentation or dissociation of protein assemblies. Collisions with background inert small gas molecules (He, N_2, Ar, Xe) are widely used for this purpose, as this approach is characterized by its ability for a slow and gradual buildup of internal energy[85,189,190]. In particular, for native MS of large macromolecules, desolvation can be promoted by using heavier gas molecules (*e.g.*, Kr, Xe), as they generally deposit more energy into the precursor molecule with each collision[85]. Since ion activation by CID is m/z and charge-state-dependent, the ions with lower charge within a charge state envelope of a particular precursor generally display substantially broader peaks due to unequal levels of desolvation[34]. Particularly large protein assemblies measured at high m/z can therefore benefit from the development of alternative desolvation methods that are less charge-state-dependent. Recent work demonstrated that in particular IRMPD could also be used to enhance desolvation, possibly even more effectively than CID[191-193], and should therefore possibly be more explored for native MS[194].

Effective desolvation in native MS is a balancing act in removing unwanted adducts while preventing the loss of specific interactions and labile PTMs. In native MS of membrane proteins, for example, sufficient desolvation is essential for stripping proteins of detergent molecules or other solubilizing agents while retaining specifically bound lipids or other small molecules (*e.g.*, drugs)[195,196]. On the other hand, there are also protein assemblies that require small molecules or specific cations to stabilize their native structures[197]. Furthermore, labile PTMs that are often present on native protein assemblies, *e.g.*, glycosylation and phosphorylation, are known to be susceptible to dissociation upon harsher desolvation conditions. Thus, in native MS, optimal desolvation must always be balanced to avoid inducing fragmentation or the loss of small molecules or labile PTMs that are essential for the structure and stability of the studied protein complexes. As each protein complex exhibits its own biochemical and biophysical features, there are unfortunately no general rules to determine the best way to achieve this.

6.2. Revealing Structural Features of Protein Assemblies by Gas-Phase Ion Activation

Perhaps the most intriguing benefit of ion activation in native MS is its ability to provide information about the quaternary structures of protein assemblies. In tandem MS experiments, the precursor ions are mass-selected and subsequently activated, which may lead to the specific ejection of individual subunits[86,198] or other modes of complex partitioning[187,199,200]. The resulting mass measurement of the monomeric dissociation products and corresponding concomitant higher-mass dissociation products can reveal the compositional buildup of the precursor. Often the order of subunit ejection and the observed dissociation pathways can provide information on the arrangement of subunits within the protein complex[198]. Moreover, additional fragmentation of the ejected monomeric subunits can be used to reveal the identity of the subunits. However, from all of the available tandem MS studies on protein assemblies, it also has become apparent that the smaller peripheral subunits often display a preference to be ejected and also that the structures of the activated protein complexes can drastically change when transferred into the gas-phase and even more so upon activation[201]. To summarize, although it should be used with care, tandem MS studies can provide essential structural information that is not easily accessible by other means, as briefly illustrated below with a few prominent examples.

6.2.1 Collisional Activation in the Gas-Phase

Collisional activation in the gas-phase has been used most extensively in native MS for protein complex dissociation, and it is featured by a wide range of instruments. Recently, Wang *et al*.[201] demonstrated that, although collisional dissociation (CID/CAD/HCD) preferentially results in ejection of peripheral subunits, less exposed nonperipheral subunits also can be ejected from certain protein assemblies. Studying the 20S core proteasome of *Thermoplasma acidophilum*, engineered antibody complexes, and elongated complement protein complexes, they identified two major pathways by which nonperipheral subunits can be released. First, sequential dissociation of highly charged peripheral subunits may transfer residual energy back to neighboring subunits and increase their exposure. Alternatively, nonperipheral subunits may be ejected either directly from complexes stabilized by chemical charge reduction or from elongated complexes that may undergo compaction upon transfer into the gas phase. Several other examples of studies that combine native and tandem MS using CID/CAD/HCD will be described in more detail below, particularly in **section 7**.

Generally, the mechanism of collisional dissociation of protein complexes involves the release of weakly bound single subunits that take away a large portion of charges carried by the precursor. Several groups have tried to come up with an explanation for this frequently observed behavior, which is known as asymmetric charge partitioning. In addition to available empirical evidence, the Konermann group performed molecular dynamics simulations that provide a more detailed fundamental understanding of how multimeric complexes behave upon collisional dissociation[202]. In brief, because of the slow build-up of internal energy inherent to collisional activation methods, extensive structural

rearrangements and unfolding can take place prior to dissociation[203-206]. This is attributed to a microsecond-time-scale process in which the to-be-ejected single subunit unfolds, leading to Coulombically favored charge redistribution. Consequently, the high-charge monomer subunit departs from the concomitantly formed charge-stripped high-mass fragment[206,207]. Thus, although widely used, protein complex partitioning in collision-based native tandem MS is a complicated process, and further studies are needed to understand abnormal dissociation pathways observed for complex heteromeric assemblies.

6.2.2. Surface-Induced Dissociation

SID combined with native MS has been pioneered primarily by the group of Wysocki as an alternative approach that, in contrast to dissociation methods based on collisions with gas molecules, largely avoids unfolding of individual subunits and asymmetric charge partitioning[199,200,208-210]. Collision with a surface is a relatively fast activation process that provides access to higher-energy dissociation pathways and occurs on the picosecond time scale, and it can lead to the direct ejection of folded subunits even from stably bound assemblies resistant to CID[203,211]. Such a mode of partitioning is mostly accompanied by symmetric charge redistribution between the dissociation products, which allows SID to retain a high level of structural information that, when combined with IM-MS, is sensitive enough to reveal differences between precursor conformations[212,213]. Moreover, ejected subunits largely preserve non-covalent interactors because of their folded nature, enabling SID to probe stoichiometries of ligand binding. In a study of structurally and energetically well-characterized protein complexes, it was shown that the strength of the interaction interface in a given protein assembly could be correlated with the amount of energy required for its dissociation by SID[214]. Illustratively, Vimer *et al*.[215] have recently utilized these approaches to probe structural conservation and variation among orthologous 20S proteasome complexes purified from five different organisms (described in more detail in **section 7.2**). Structural information obtained from SID and covalent labelling studies have also enabled Song *et al*.[216] to model and refine the structure for the hexameric toyocamycin nitrile hydratase.

6.2.3. Ultraviolet Photodissociation

UVPD was initially exploited in mass spectrometry for applications in bottom-up and top-down fragmentation[184,187,217-219]. Its applicability to dissociate native protein complexes has more recently been explored. Using high-energy 193 nm photons, UVPD is able to deposit substantial amounts of energy on a time scale close to that of SID, giving access to higher-energy dissociation pathways[220-222]. Two recent studies have shown that high-energy laser pulses (≥ 1.5 mJ) can eject subunits from protein complexes with more symmetrical charge partitioning, in some cases resembling SID-like behavior[223,224]. Contrary to SID, however, UVPD can provide additional backbone fragment ions, yielding sufficient sequence coverage for the identification of individual subunits[225]. The efficiency of UVPD also exhibits little dependence on the precursor charge compared with collision-

based approaches[226], making this approach well-suited also to study large native protein assemblies that are detected in multiple charge states. As an example, this allowed Greisch *et al.*[123] to eject intact 16 kDa subunits from a 1 MDa AaLS virus-like particle while simultaneously providing diagnostic fragment ions that covered around 60% of the subunit's sequence[28].

6.2.4 Infrared Multiphoton Photodissociation

The use of powerful infrared lasers for the fragmentation of proteins and protein assemblies has also been explored in native MS. IRMPD, although very similar to collisional dissociation, has several advantages. First, it does not require elevated gas pressures, which is especially important in FT-ICR and Orbitraps that need a high vacuum for optimal performance. In addition, IRMPD provides a high level of control over the amount of energy deposited by photon irradiation. These advantages prompted researchers to implement IRMPD on the various instruments used for native MS, including Q-ToF[227], FT-ICR[6] and Orbitrap-based instruments[194]. The groups of Loo[28] and Gross[228] have extensively explored the application of IRMPD for native top-down MS on ICR platforms, primarily using this method for supplemental activation of proteins and protein complexes along with fragmentation using more energetic and prompt activation techniques like ECD. In an attempt to study membrane protein complexes Mikhailov *et al.*[227] implemented IRMPD on a Q-ToF instrument, whereby IR irradiation provided great results for liberating non-soluble protein assemblies from detergent micelles. Finally, in a very recent study combining IRMPD with native MS analysis, Greisch *et al.*[194] explored the dissociation of protein complexes with masses up to 1 MDa, revealing that the increasing cross sections of larger protein assemblies enable more efficient irradiation and fragmentation.

6.2.5. Electron-Based Dissociation Techniques

Electron-based fragmentation techniques include a range of methods that vary in the way electrons are delivered to the proteins. For instance, in electron-capture dissociation (ECD) the analyte is directly irradiated with electrons, while in electron-transfer dissociation (ETD) the fragmentation is induced via ion-ion interactions between positively charged protein ions and the negatively charged electron-transferring reagent anions. Because electron-based activation on its own does not fully induce disruption of non-covalent interactions, observed fragments primarily stem from the most exposed regions on the protein surface. Thus, for high sequence coverages, it is often necessary to use supplemental activation to release the formed fragments. By using ETD on a Q-ToF instrument for fragmentation of protein complexes, Lermyte and Sobott[229] have established that, by balancing the supplemented activation it is possible to probe surface-exposed regions and gain insights into protein behavior upon collision-induced unfolding. Leveraging the ability of ETD to preserve labile modifications, Tamara *et al.*[230] applied ETD in MS^3 on an Orbitrap Lumos instrument to position phosphate transfer events within the binding interface between Pin1 and its phosphorylated binding partner. ECD implemented

on an FT-ICR mass spectrometer was successfully applied for structural investigation of various protein assemblies, exposing flexible regions[228] and binding interfaces in protein-ligand complexes[231]. Native ECD, which is induced by the generation of endogenous electrons in the areas adjacent to metal-binding sites through infrared excitation, allowed researchers to map iron-binding pockets in a ferritin complex[232]. Using yet another electron-based activation method, electron ionization dissociation (EID), in their studies of metal-binding complexes, Li *et al*.[233] could reflect structural differences between apo, Zn-, and Cu,Zn-SOD1 dimeric complexes and probe the structural stability of carbonic anhydrase I.

6.3. Higher-Order MS^n Methods in Native MS

Characterization of proteoforms and subcomplexes can be achieved by pseudo-MS^3 analysis, whereby the protein complexes are interrogated in a two-step process[118]. Such experiments have been implemented on Orbitrap UHMR instruments and have shown powerful capabilities in providing structural information even when collisional activation is used in both steps[70]. For example, van de Waterbeemd *et al*.[52] employed collisional dissociation to investigate the stalk complexes in chloroplastic ribosomes. First, these subcomplexes were ejected from intact ribosomal particles by collisional activation in the front end of the mass spectrometer and then dissociated further with HCD to determine their exact stoichiometry. By combining complementary fragmentation methods in two-tier ion activation on an Orbitrap UHMR, first using collisional activation in the front-end of the instrument and then UVPD in the back-end, Mehaffey *et al*. were able to sequence ejected subunits and position the non-covalently bound cofactors on human mitochondrial BCAT2[122]. In another approach, Li *et al*.[28] combined native MS with top-down sequencing on a high-resolution FT-ICR instrument. They demonstrated that assemblies with masses of up to 1.8 MDa can be resolved in charge states similar to those observed in ToF or Orbitrap experiments. Next, they employed various ion activation techniques to investigate the compositional makeup of various protein complexes. In-source dissociation (ISD) or CAD was used to eject individual subunits, followed by IRMPD to reveal information on the backbone sequence and PTMs, thereby characterizing proteoforms in various assemblies, including β-galactosidase. On the other hand, ECD fragmentation was used to investigate the topology of the intact complex. This activation method largely preserves non-covalent interactions, meaning that when N- and C-terminal fragments are observed, they likely originate from the surface of the complex. The integrated platform presented by Li *et al*. thus shows great promise for native MS as a bridge between proteomics and structural biology, providing new insights into protein structure and function. Higher-order tandem MS methods have also been used to study the interactions of membrane proteins with endogenous lipids, peptides, and small molecules. In such an application, Ro *et al*.[234] interrogated the membranous metalloenzyme pMMO embedded into nanodiscs (**Figure 9**). In the first collisional activation step, they released the complex from a solubilizing nanodisc or, by elevating collisional activation energies, even dissociated the complex into distinct pMMO subunits. Subsequently, the ejected subunits were further fragmented in the HCD cell of the Orbitrap

instrument. Overall, this approach allowed the researchers to determine the stoichiometry of copper binding and the presence of several PTMs in the pMMO complex.

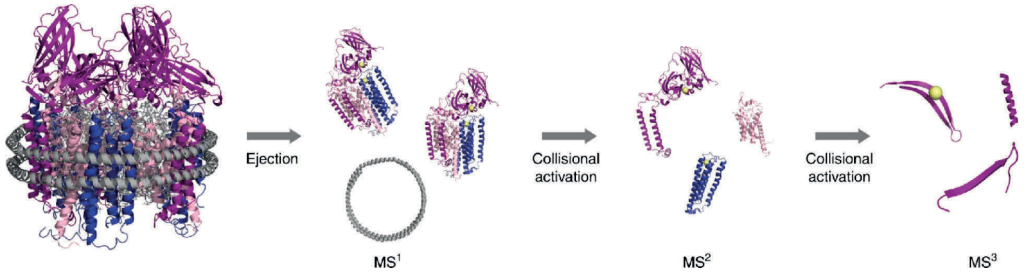


Figure 9 | Example of a higher-order native top-down MS approach used for the characterization of membrane proteins such as pMMO. Nanodiscs embedded with pMMO were subjected to CID in the source region for ejection of the protein complex from the nanodisc, resulting in the stabilization of the pMMO protomer species (MS^1). Increasing collisional activation breaks the protomer up into individual pMMO subunits (MS^2). Further collisional activation in the HCD cell of the Orbitrap UHMR enables backbone fragmentation of each subunit (MS3). From ref [234]. CC BY 4.0.

Recently, the Robinson group introduced the concept of "nativeomics" as a multilevel native top-down MS approach to identify endogenous ligands of membrane protein assemblies, distinguishing them from other copurified lipids[75,235]. The first step of the workflow introduced by Gupta *et al.*[235] involves stepwise delipidation of the protein assembly, using preservation of the oligomeric state as a measure for conditions that still allow key ligands to remain bound. The extracted lipid fractions are then subjected to LC-MS analysis, providing a library of potential ligands. High-energy native MS measurements were performed on a ToF instrument modified for dissociation of detergent micelles in the front end of the instrument. This enabled mass selection of a single charge state of the protein-ligand complex followed by dissociation of the ligand in the collision cell to determine its intact mass. Database searches of the mass uncovered in this manner were then performed to determine the identity of the ligand. In a follow-up study, Gault *et al.*[75] utilized a modified Orbitrap instrument to enable systematic selection and fragmentation (up to pseudo-MS^6), providing further confidence in ligand assignments and expanding the approach to more complex samples. Studying endogenous lipids binding to aquaporin Z (AqpZ), they identified interactions with various lipid families, with AqpZ showing a preference for shorter and unsaturated chains. Similarly, they were able to identify previously unknown ligands bound to dimeric outer mitochondrial membrane translocator protein (TSPO, **Figure 10**), demonstrating the utility of tandem MS for identification of potential regulators of this critical drug target.

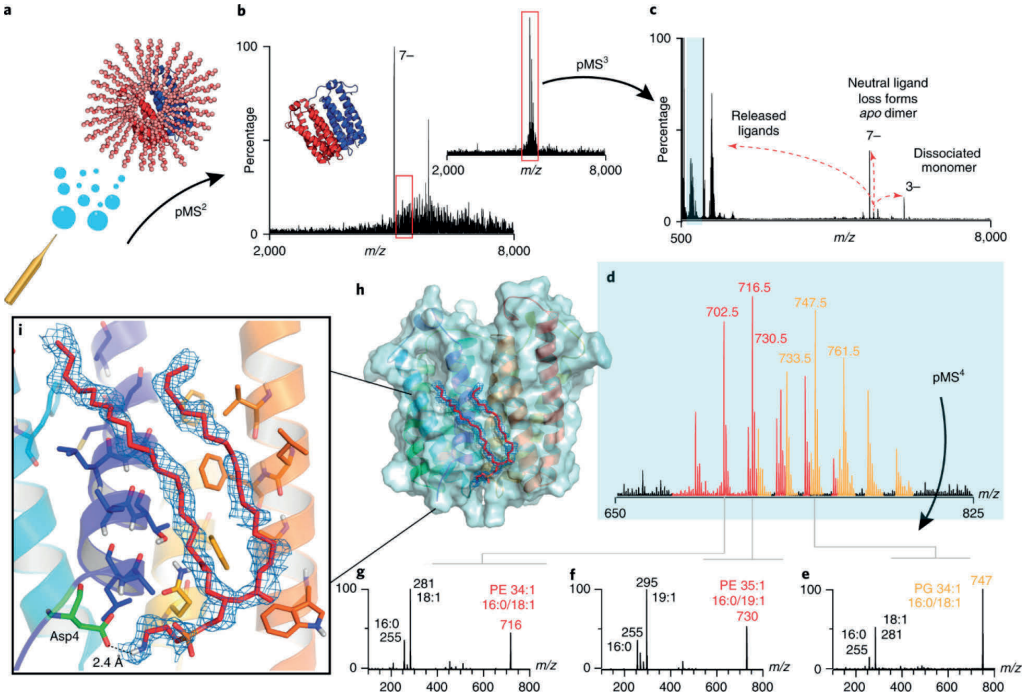


Figure 10 | Identification of unknown ligands bound to TSPO and subsequent fitting of PE 16:0/18:1 lipids into unresolved electron density in the X-ray structure. (a) Schematic showing the release of the TSPO dimer from detergent micelles (the subunits are shown as red and blue cartoons and the detergent micelle as orange spheres). **(b)** Native mass spectrum of the TSPO dimer (pseudo-MS^2 step (pMS^2)) and isolation of the 7- charge state (red box) (inset). **(c)** Collisional activation (pseudo-MS^3 step (pMS^3)) yields dissociated monomer, apo dimer produced from neutral ligand loss, and multiple ligands at low *m/z* (blue box). **(d)** Zoom of the low-*m/z* region showing two peak series corresponding to multiple lipids: PE (red) and PG (orange). **(e-g)** Isolation and subsequent fragmentation of released lipids (pseudo-MS^4 step (pMS^4)) define the hydrocarbon chain length and extent of unsaturation. **(h)** Position of the most abundant PE lipid identified, PE (16:0/18:1) (red sticks), in the TSPO A139T structure (PDB entry 4UC1). **(i)** Critical protein-lipid interactions (zoom box). Reproduced with permission from ref [75]. Copyright 2020 Gault *et al*.

6.4. Combining Tandem MS with Native Separation in Native Top-Down Proteomics

When samples contain multiple protein assemblies with similar masses, it is quite challenging to characterize the constituent assemblies with direct-infusion native MS, and therefore, a separation step, akin to liquid chromatography used in shotgun proteomics, is required. Native top-down proteomics is a relatively new direction in the field that aims to analyze proteomes at the level of endogenous protein assemblies[236]. This approach utilizes a mild (native) separation technique such as size-exclusion chromatography (SEC), ion-exchange chromatography (IEX), or capillary electrophoresis (CE) to first fractionate complex biological samples, followed by native tandem MS to analyze intact protein assemblies and their constituents[171,237,238]. In a recent study by Skinner *et al.*[171], a two-step off-line fractionation method was used to characterize 125 protein assemblies

and 217 distinct proteoforms from human cell lines and mouse tissue. An alternative approach by Shen *et al*.[237] combined SEC prefractionation with online capillary zone electrophoresis (CZE), providing a high-throughput platform that identified 144 proteins, 672 proteoforms, and 23 protein complexes from *Escherichia coli*.

In **section 6**, we summarized how native MS can be combined with tandem mass spectrometry and how this combination can help to improve the desolvation of the produced ions, resulting in more precise mass measurements. Tandem MS can also be used to gain insights into the quaternary structures of the studied protein assemblies. The latter significantly benefits from the monitoring of distinctive dissociation pathways, which provide information about the connectivity of loosely bound peripheral subunits, subunit topologies, proteoform identities, and sometimes even cofactor binding. More examples of the combination of native MS and tandem MS will be described later in the review, especially **section 7**, wherein we highlight recent work on different classes of proteins and protein complexes as studied by high-resolution native MS.

7. Application of High-Resolution Native MS in the Characterization of Protein Assemblies

As an analytical technique, native MS has been around now for several decades. Although it was initially used to study small proteins and protein-ligand interactions, the applications of native MS have expanded and by now cover a huge variety of protein assemblies involved in a plethora of different biological processes. Although we aim to provide an extensive review, we cannot be all-inclusive in describing all of the reported applications of high-resolution native MS in the characterization of protein assemblies. Instead, we provide and review a selection of applications categorized by the biological processes in which these assemblies are involved, notably protein synthesis and degradation, complement activation, and light harvesting. Additionally, we review the application of native MS to characterize virus(-like) particles and membrane-embedded protein assemblies. The focus and scope of this review necessitate that we leave out some other exciting applications, so we apologize especially to the authors of these studies for not being fully comprehensive in this section.

7.1. Protein Assemblies Involved in Protein Synthesis: Ribosomal Particles

Ribosomes are very large macromolecular machines found within all living cells that are responsible for protein synthesis (*i.e.*, mRNA translation). Ribosomes consist of two major components: the small and large ribosomal subunits. Each subunit consists of one or more large ribosomal RNA (rRNA) molecules and many ribosomal proteins (RPs). As such, ribosomes represent rather challenging samples for native MS. First, they are very large and multicomponent ribonucleoprotein assemblies that display multiple putative PTMs and composition-related heterogeneities. Additionally, significantly negatively charged

nucleic acids, originating from the rRNA, result in a decreased total positive charge of the ribosomal particles, leading to detection at elevated m/z compared with merely proteinaceous assemblies of similar mass. Perhaps because of these aspects, the application of native MS to the analysis of ribosomes has been somewhat limited. Structural analysis of ribosomal particles has been primarily achieved with crystallography and now more and more by cryogenic electron microscopy (cryo-EM), sometimes supplemented by chemical cross-linking mass spectrometry[239-241]. Some novel detailed insights into ribosome composition can also be obtained via bottom-up mass spectrometry[242,243]. Recently, the earlier view that the ribosome is a static system has been challenged on the basis of multiple lines of evidence for a more heterogeneous and dynamic nature that likely has implications for translation and protein synthesis[244]. The latest studies have shown that the current state of high-resolution native MS can provide a high level of insight into compositional and structural heterogeneity of ribosomal particles, especially when combined with other modes of mass spectrometry or other structural biology tools[52,65].

In the early 2000s, the limited mass resolution and desolvation capabilities of native-MS-capable mass spectrometers hampered accurate mass detection, leading to blindness toward the substoichiometric or low-abundance variants of ribosomal particles[245,246]. However, the Robinson group demonstrated the potential of native MS to provide useful information on ribosome composition by unraveling the stoichiometry and proteoforms of the stalk complex within the 50S ribosomal particle derived from *Thermus thremophilus*[245]. These findings complemented more standard high-resolution structural techniques such as crystallography, which were limited in providing high-resolution structural information for the more flexible parts of the ribosomes. Although they were unable to obtain accurate masses of ribosome assemblies because of large peak widths and low overall resolution, in 2006 McKay *et al*. reported that the major bottlenecks for solving this issue were caused by incomplete desolvation of the produced ions and potentially co-occurring components within ribosomal complexes[247]. Notably, by modeling the average mass increase as a function of peak width for standard non-covalent assemblies, they could better explain the observed mass distributions of the ribosomal particles. Tandem MS experiments further confirmed that multiple ion populations did coexist in their experiments.

About 10 years later, the advanced desolvation and transmission capabilities of Orbitrap instruments with ultrahigh mass range (Orbitrap UHMR)[30] have enabled the recoding of more detailed high-resolution mass spectra, revealing altered ribosomal particles and substoichiometric subunits (**Figure 11**). Such spectra were achieved mainly as a result of improved instrumental sensitivity and desolvation in both the front and back ends of the instrument and by the use of charge-reducing conditions, *i.e.*, addition of triethylammonium acetate to the spraying solvent[65]. The mass accuracy for detection of the *E. coli* 70S ribosome was within 0.7% of the theoretical mass, an excellent result for a particle with a mass of more than 2 MDa. By lowering the concentration of Mg^{2+} ions essential for 70S complexation, smaller and better-resolved spectra of the 30S and 50S ribosomes could be recorded, revealing their intrinsic heterogeneities, which partly also explained the relatively less resolved spectra of the intact 70S ribosome.

z

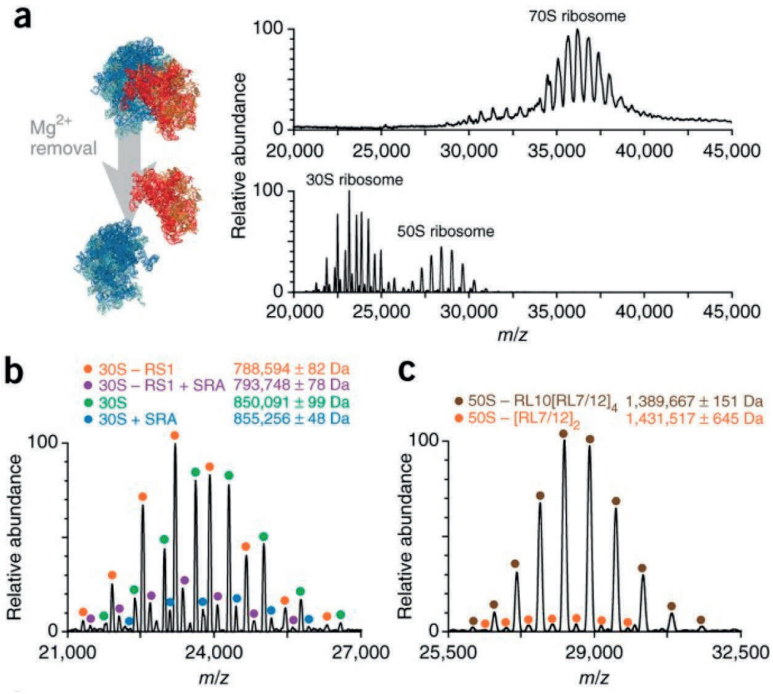


Figure 11 | High-resolution native MS of ribosomal particles. (a) (left) Structures of the *E. coli* 70S ribosome consisting of the 50S (proteins shown in blue and rRNA in green) and 30S (proteins shown in red and rRNA in orange) particles. (right) High-resolution native mass spectra of the 70S, 30S, and 50S ribosomes. **(b)** Distinct particles of the 30S ribosome, with annotations provided at the top. Masses are shown as mean ± SD. **(c)** Distinct particles of the 50S ribosome, with annotations provided at the top. In most of the particles, the pentameric stalk complex is absent. Reproduced with permission from ref [65]. Copyright 2017 Nature Publishing Group.

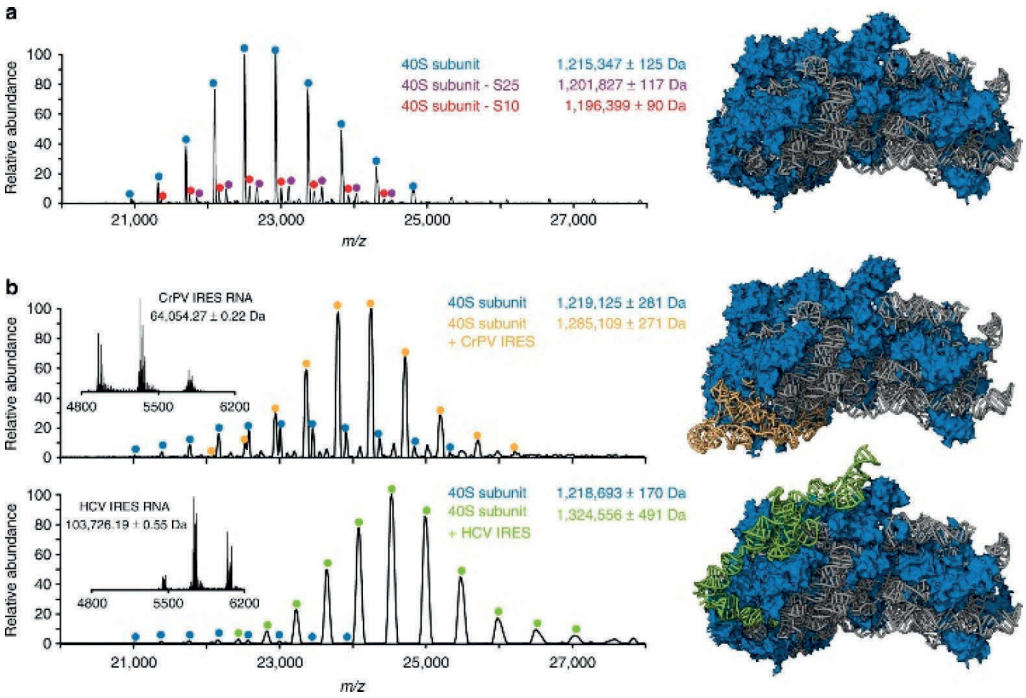


Figure 12 | High-resolution native mass spectra of free and internal ribosome entry site (IRES) RNA-bound human 40S subunits. (a) Native mass spectrum of the human 40S subunit acquired with an Orbitrap UHMR mass analyzer. The well-resolved charge states of three distinct forms of the ribosomal subunit could be detected. The most abundant fully assembled 1.2 MDa 40S particles are labeled in blue, while lower-abundance particles lacking either the S25 or S10 protein are labeled in magenta and red, respectively. **(b)** Monitoring of the formation of a complex containing human 40S ribosomes and IRES RNA fragments of cricket paralysis virus (CrPV) and hepatits C virus (HCV). The mass spectra of the RNA fragments alone (insets) provide the accurate masses of the IRES elements. Structures of the free 40S ribosomes (PDB entry 5A2Q) and particles bound by CrPV (PDB entry 4V91) and HCV IRES (PDB entry 5A2Q) are shown, with the ribosomal proteins in blue, the rRNA in gray, and the IRES elements in yellow and green, respectively. From ref [52]. CC BY 4.0.

In a recent study, Abdillahi *et al.*[248] applied anion-cation reactions to study the attachment of high-mass myoglobin ions $[hMb - 11H]^{11-}$ to positively charged 30S ribosomal particles prepared by native MS. They demonstrated that these ion-ion reactions can facilitate obtaining well-separated ion signals that can yield confident charge-state and mass assignments from otherwise poorly resolved signals. Only upon attachment of multiple myoglobin anions could the intact 30S particles and the particles missing the S1 protein be resolved. This is a promising new development that is somewhat similar to the more well-established proton transfer reaction (PTR)[249], and especially can help to disentangle poorly resolved ion signals on instruments that have inherently lower resolution.

7.2. Protein Assemblies Involved in Protein Degradation: Proteasomal Particles

Proteasomes are ubiquitously present in all eukaryotes and archaea and also in some bacteria, where they regulate the abundance of particular proteins and remove misfolded or damaged proteins by degradation. This degradation process yields peptides, which can then be further degraded for use in the synthesis of new proteins or presented to immune cells by major histocompatibility complexes (MHCs) on the cell surface. The full 26S proteasome holoenzyme is built from two subcomplexes, the 20S catalytic core particle and one or two 19S regulatory particles that recruit, select, and unfold ubiquitinated substrates, preparing them for degradation in the 20S catalytic core particle. The 26S complex is a multimeric assembly with a mass of approximately 1.5 MDa, whereas the 30S complex contains the 20S core with two 19S lids. In the case of the 26S proteasome, progress in elucidating its structure has for a long time been hampered by the complexity of the system and its variability and fragility. Through advances in cryo-EM, several structures of the full 26S proteasomal particles are nowadays available[250-252]. As far as we know, no native mass spectra have been reported to date for the full 26S proteasome, but several groups have been working on the 20S core particle and the 19S regulatory lid.

The overall structure of the 20S core proteasome is highly conserved, forming an ~ 700 kDa barrel-shaped compartment whose proteolytic active sites are restricted to its interior so that only proteins entering the chamber are degraded. The 20S proteasome is composed of 28 subunits, arranged in a cylindrical structure consisting of four heptameric rings: two outer α-type subunit rings embracing two central β-type subunit rings ($\alpha_7\beta_7\beta_7\alpha_7$). The two outer α rings function as a gate that regulates the entry into the proteolytic chamber confined by the two β rings. Although the 20S proteasome is ubiquitously present in archaea and eukaryotes, a substantial increase in complexity and diversification of the complex is observed with the evolution of different organisms. Prokaryotic 20S proteasomes (*e.g.*, from *Thermoplasma acidophilum*) are generally composed of 14 identical α subunits and 14 β subunits, while in eukaryotic proteasomes, the α and β subunits each differentiate into seven distinct subtypes, accounting for a total of 14 unique subunits. Moreover, in mammalian systems, dedicated distinct proteasome particles co-occur next to the standard (or constitutive) proteasome, such as the immunoproteasome (which contains distinctive βi subunits), the thymoproteasome, or the testis-specific proteasome.

Early in the era of native MS, several studies reported spectra of 20S proteasome core particles originating from different species. These studies readily provided highly charge-resolved spectra for archaeal and yeast 20S proteasomes, but the native MS spectra became more cluttered in the case of more heterogeneous particles originating from mammalian cells[18,27,253,254]. These initial studies were often combined with tandem mass spectrometry to define the masses of the subunits. Later studies, apart from structural analysis of the main proteasome assemblies, also focused on the binding and stoichiometries of the different proteasomal inhibitors and substrates[253].

A large body of the work performed by native MS to study the 20S proteasome has recently been reviewed by Ben-Nissan *et al.*[255]. They described several studies on how native MS

was used to reveal the composition, intrinsic heterogeneity, stoichiometry, subunit architecture, and topology of the 20S particles. Also, they provided protocols on how to extract endogenous 20S proteasomes from yeast, rat liver, and human cells that are amenable for analysis by native MS. In a subsequent elegant study, Vimer *et al.*[215] used these enrichment protocols for the purification and subsequent analysis by native MS combined with IM-MS and top-down MS to probe structural and functional conservation between 20S proteasome particles extracted from archaeal *Thermoplasma acidophilum*, yeast (*Saccharomyces cerevisiae*), and mammals, including rat (*Rattus norvegicus)*, rabbit (*Oryctolagus cuniculus*), and human (*i.e.*, from HEK293 cells) (**Figure 13**). By using native IM-MS, they observed higher collision cross sections (CCSs) for the eukaryotic proteasomes compared with those of the archaeal 20S complex, which they related to the increased complexity of eukaryotic 20S particles. Distinctively, the eukaryotic 20S proteasome particles exhibited several PTMs, extending their diversification. As these latter proteoform variabilities could not be directly resolved from the native MS spectra alone, their elucidation required the combination of native MS with top-down proteomics using both HCD and ECD as fragmentation methods.

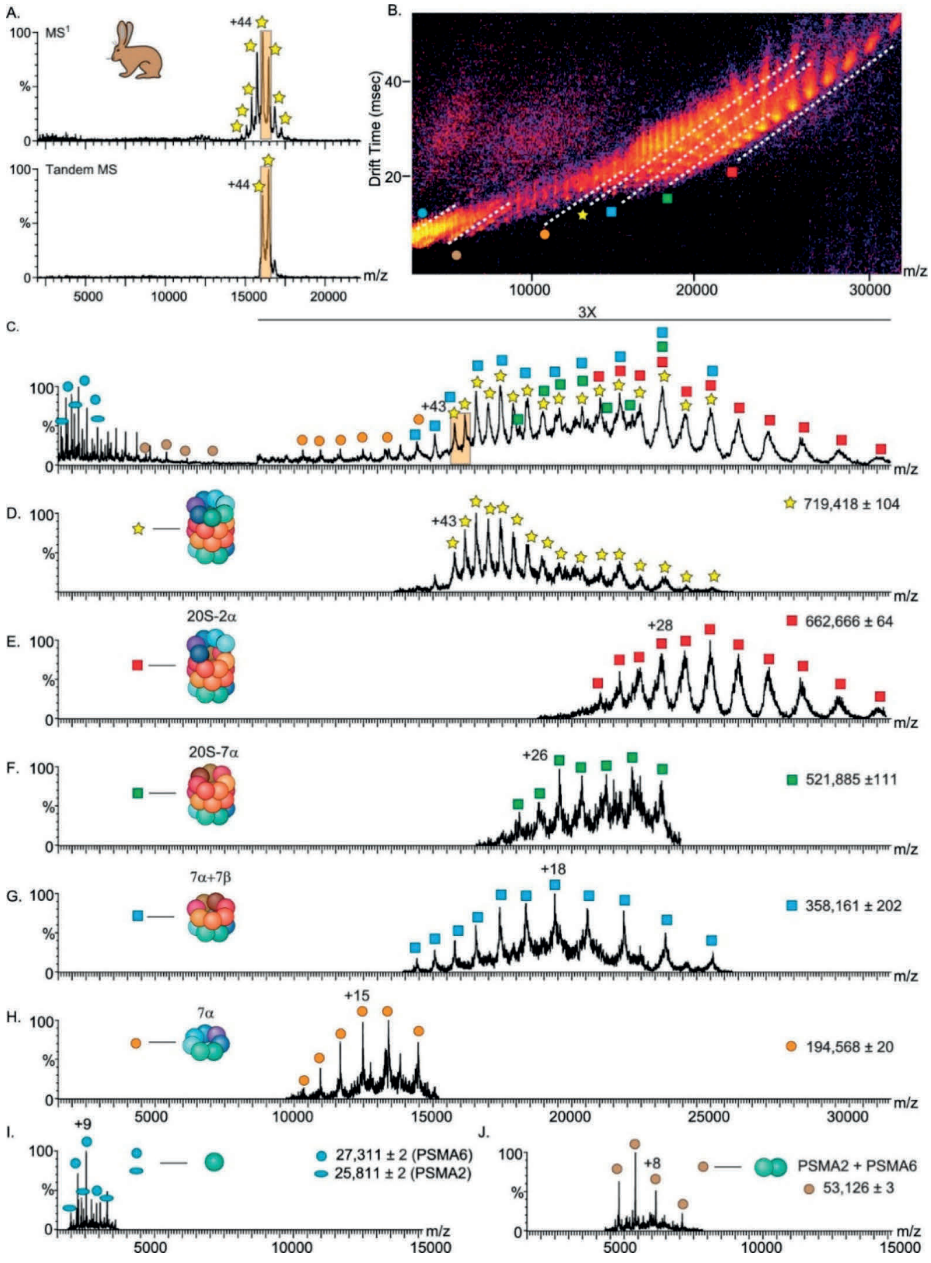


Figure 13 | 20S proteasomal particles analyzed by surface-induced dissociation (SID) combined with ion mobility mass spectrometry (IM-MS). SID-IM-MS spectra of the rabbit 20S proteasome complex reflect the cylindrical topology of the complex. **(A)** Charge-reduced rabbit 20S proteasome precursor ions were measured on a Synapt G2 instrument equipped with an SID cell (top), and the 43+ and 44+ charge states were isolated (bottom) and accelerated into the surface at 150 V. **(B)** IM-MS plot of the SID spectra of the rabbit 20S proteasome. The separation in drift time (vertical axis) assists in the discrimination of species that overlap in *m/z* (horizontal axis), as shown in **(C)**. The major populations of the dissociation products are designated by dashed lines in (B) and labeled with symbols that are graphically depicted in **(D-J)**. The extracted *m/z* spectra from the underlined regions in the IM-MS plot (D-J), show the identified dissociation products. Reproduced from ref [215]. Copyright 2020 American Chemical Society.

Although a large body of reported work has been focused on the 20S core particles, Sharon *et al.*[256] did report on analyzing the structural organization of the yeast 19S proteasomal lid by native MS. Their native MS data on the intact lid complex demonstrated that eight of the nine subunits were present stoichiometrically (with the Rpn6 subunit missing), and they observed a stable tetrameric subcomplex. Tandem MS was used to infer details on the subunit architecture at a time when little structural information was available for the lid.

A challenge in the analysis of proteasomes is to detect the 26S and 30S proteasome complexes by native MS, which has remained difficult because of the instability of these holo enzymes. Using mass photometry in parallel with native MS may be of assistance here. In this context, Sonn-Segev *et al.*[257] studied the stability of the proteasome purified from bovine heart tissue by mass photometry and observed distinct 2.4 MDa 30S particles (two regulatory particles and one core particle) and 1.5 MDa 26S particles (one regulatory particle and one core particle) as well as the 700 kDa core particle and 800 kDa regulatory particle. In line with earlier literature, their data reveal that the holo enzyme dissociates at ionic strengths above 50 mM. Screening sample stability under varying conditions by mass photometry may help to interrogate distinct proteasomal assembly states with native MS.

7.3. Virus(-like) Particles

Even in the early years of native MS, because of their biophysical stability, viruses and virus-like particles (VLPs) became targets for mass analysis, primarily to probe whether such large assemblies could be transferred intact into the mass spectrometer and subsequently mass-analyzed[258,259]. For detection, either early-generation charge-detection-based devices[259], electrophoretic mobility (*i.e.*, GEMMA)[260-262] or Q-ToF mass analyzers were employed, with the latter allowing the most accurate mass analysis via (partly) resolved charge states[141]. Some of the first high-resolution data were obtained for intact capsids of hepatitis B virus (HBV) by making use of a Q-ToF mass analyzer modified for high-mass analysis[87,263]. HBV capsids are quite unique in being able to exhibit two distinct icosahedral morphologies in vivo, composed of 90 or 120 dimers with masses of roughly 3 and 4 MDa, respectively. These early native mass spectra displayed well-separated charge state distributions for both types of morphologies, enabling mass assignments within 0.1% error. The HBV capsids were surprisingly stable during transfer into the gas phase and through the vacuum of the mass spectrometer. Measuring the CCS of the capsids by ion mobility mass spectrometry allowed an estimation of the capsid radii that was in good agreement with the particle dimensions as observed by EM, suggesting a largely retained native capsid morphology in the gas phase[263].

Since this early work, native MS has contributed substantially to the field of structural virology, particularly in regard to our understanding of capsid assembly, virus maturation, genome packaging, viral protein subunit stoichiometries, and virus stability. This work has been summarized in several excellent review articles, to which we refer readers with a deeper interest in mass-spectrometry-based structural virology[127,264-267]. We divide this subsection into applications focusing on empty viral capsids, endogenous genome-packed

viruses, and virus-like assemblies, including bacterial compartments and natural as well as designed synthetic protein nanocages. The breadth of these applications already demonstrates that high-resolution native MS has become an integral biophysical tool for studying these viral assemblies.

7.3.1. Empty Capsids and Virus-like Particles

When a virus capsid is built up from a single type of subunit that carries no PTMs and is not affected by truncations due to processing, the resulting native mass spectra for the empty capsid can often be nicely charge-resolved by native MS, even when the whole assembly has a molecular weight of several megadaltons. This has been demonstrated in recent years for a variety of viral assemblies such as the capsid of MS2[141], the hepatitis B virus capsid[65,87], norovirus subparticles[89,268], portal oligomers of the phages P22, Phi-29, and SPP1[269], and the 18 MDa bacteriophage HK97[34,270].

Many virus capsids are built up from multiple distinct subunits with unique intrinsic structures, sequences, and masses. A quite interesting example of that is the adeno-associated virus (AAV), which nowadays is receiving considerable attention because it is one of the most explored viruses for gene delivery and therapy. AAVs are small (20-25 nm in diameter), non-enveloped viruses with a T = 1 icosahedral capsid built up of 12 pentamers. Thus, the AAV capsid consists of a total of 60 subunits made up of three distinct viral proteins originating from the same cap gene (VP1, VP2, and VP3), which vary only in their N-terminal sequences. All three VPs play crucial and specific roles in cell entry and transduction. The exact stoichiometry and organization within AAV capsids have been ambiguous for a long time, although generally it has been assumed that VP3 is dominant, with smaller amounts of VP1 and VP2 present as well. On the basis of gel, LC-MS, or CE-MS analysis of the VP monomers formed by denaturing the capsid, a VP1:VP2:VP3 ratio of around 5:5:50 is often quoted. However, a 5:5:50 capsid monomer ratio does not necessarily mean that each particle contains exactly those numbers of VP proteins. In contrast, using high-resolution native MS, first Snijder *et al*.[116] and more recently Wörner *et al*.[271] demonstrated that AAVs assemble quite randomly from the VP capsid proteins available in the host cell. This results in an ensemble of particles, all having a different distribution of subunits, with theoretically 1891 possible co-occurring capsid stoichiometries with different masses! High-resolution native mass spectra of intact AAV capsids (MW ≈ 3.8 MDa) displayed both highly resolved regions and regions wherein spectral interferences did occur (**Figure 14**). Through extensive spectrum simulation, Wörner *et al*. were able to resolve and annotate this spectral complexity, assessing the VP stoichiometries in a panel of AAV serotypes from different production platforms. They argued that by systematic scoring of the stochastic assembly model against experimental high-resolution native MS data, a sensitive and accurate method was created to characterize these exceptionally heterogeneous gene-delivery vectors.

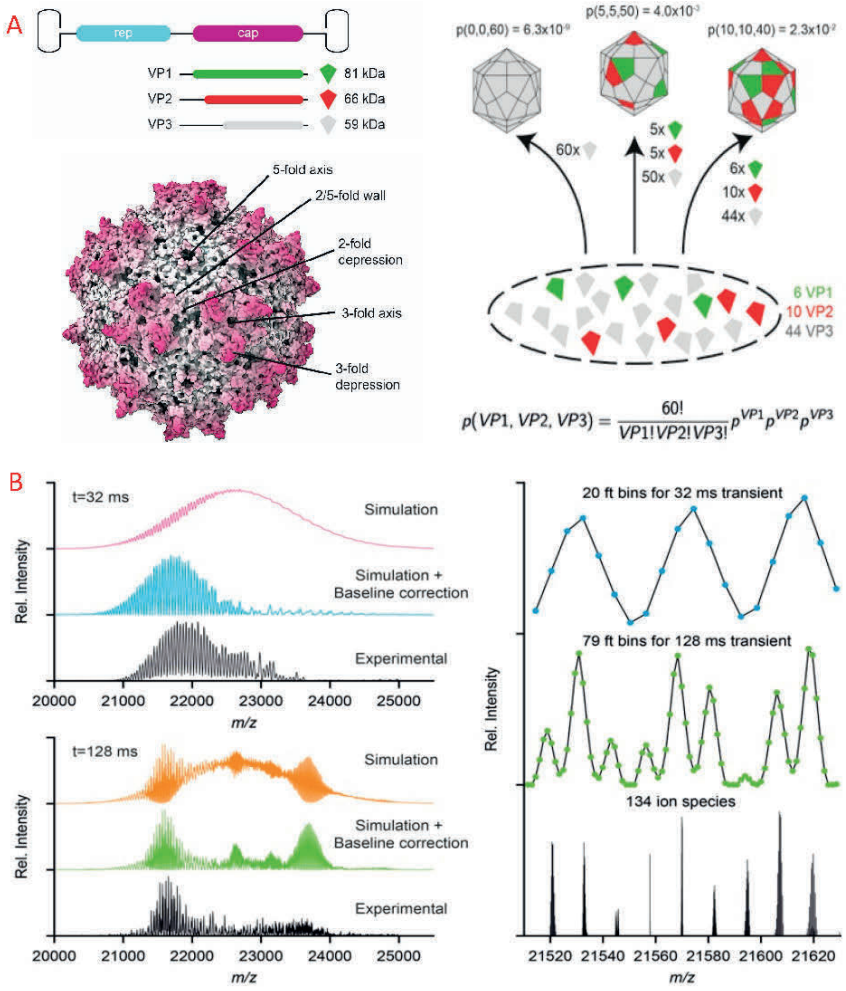


Figure 14 | Analysis of adeno-associated virus (AAV) by high-resolution native MS. **(A)** AAV (structure from PDB entry 3NG9) is composed of 60 copies of a combination of VP1, VP2, and VP3, encoded within the same cap open-reading frame, sharing a common C-terminal sequence. AAV capsid assembly follows a stochastic model, where VP proteins are taken randomly from the pool of expressed VP proteins. The probability for each stoichiometry is determined by the ratio of the expressed VP proteins. **(B)** Simulated and experimental high-resolution native mass spectra of AAV at mass resolutions corresponding to transient times of 32 and 128 ms. Each plot shows (top to bottom) the simulated mass spectra before and after baseline correction and their experimental counterparts. On the right are depicted zoomed-in parts of these spectra. Adapted from ref [271]. CC BY 4.0.

7.3.2. Genome-Loaded Authentic Viruses

Most of the high-resolution MS studies reported to date have been on empty viral capsids. Studies on genome-packed viruses are scarce, partly because analyzing such viruses typically requires special biosafety measures. Moreover, heterogeneity in the packed genome can lead to mass spectra with reduced resolving power. Consequently, most of the data reported to date have been on viruses that are not infectious for humans. The viruses

studied harbored a homogeneous RNA- or DNA-based genome. For instance, Snijder *et al*. used high-resolution native MS to investigate the structure and assembly of the picorna-like triatoma virus[272]. Interestingly, the genome-packed 8.3 MDa virus could not be charge-state-resolved, but when the genome was released, triggered by a pH jump in vitro, the empty capsid proteins reassembled into 5.4 MDa particles that could be partly charge-state-resolved using a modified Q-ToF mass analyzer. Genome loading within the plant viruses cowpea chlorotic mottle virus (CCMV) and brome mosaic virus (BMV) was studied by native MS by van de Waterbeemd *et al*., who compared data obtained with Q-ToF and Orbitrap EMR instruments[273]. Both CCMV and BMV have masses of approximately 4.6 MDa, of which approximately 1 MDa originates from the genomic content of the virion. These viruses pose an analytical challenge because they both represent mixtures of three particles carrying different genome segments that vary in mass by approximately 0.1 MDa (~ 2%). They showed that by manipulation of the charging of the particles (using additives in the electrospray spray solution), the mass spectra could be mass-resolved, enabling correct assignment of the charge states and demonstrating that native MS of these endogenous multipartite virions is possible. Although the spectra obtained by the Q-ToF and Orbitrap EMR were alike, the improved desolvation obtained by using the Orbitrap instrument generally produced better resolved mass spectra. In contrast, when CCMV was not loaded with its genome but small water-soluble phthalocyanine molecules, the mass and small-molecule content could only be estimated from a non-resolved "bump" in the mass spectra[274].

Some of the most well-resolved high-resolution native mass spectra of authentic viruses were obtained for the flock-house virus (FHV)[65]. FHV is a small, non-enveloped, icosahedral $T = 3$ virus, containing a single-stranded RNA genome comprising two genes: RNA1 (3.1 kb) and RNA2 (1.4 kb). RNA2 encodes the capsid precursor α, of which 180 copies form the initial viral capsid. Upon maturation, α undergoes an autocatalytic cleavage in its C-terminus to form β, the main structural capsid component, and a short hydrophobic g peptide required for endosome penetration that remains associated with the viral capsid but can be released at lower pH. The intact FHV virus particles, containing both strains of RNA and 180 copies of both the β and g chains, were analyzed by native MS on an Orbitrap UHMR, both under normal aqueous ammonium acetate conditions and under conditions where charging was reduced. Under both sets of conditions, clear baseline-resolved spectra could be obtained, at $m/z \approx 42{,}000$ and 57,000 Th, respectively, enabling an accurate mass determination at 9.3 MDa[65,108]. Interestingly, as these ions still carry around 200 or 160 charges, they produce sufficient image current in the Orbitrap to be detected as single ions, as demonstrated by Orbitrap-based charge detection single-particle MS[108].

7.3.3. Viral-like Assemblies, Bacterial Compartments, and Synthetic Protein Nanocages

One of the better known viral-like particles of bacterial origin is encapsulin, whose virus-like structure was originally reported by the Ban group[275]. Since then, encapsulins have been found to be present in multiple microorganisms[276]. The encapsulins from

Thermotoga maritima and *Brevibacterium linens* are formed from 60 copies of the same monomer that assemble into an icosahedral shell with a diameter of around 25 nm. They naturally encapsulate the enzyme dye-decolorizing peroxidase (DyP). DyP itself is assembled as a trimer of dimers, *i.e.*, a 240 kDa hexamer. The specific encapsulation mechanism of DyP is mediated by its C-terminal end, which interacts with a defined region of the encapsulin inner surface. Snijder *et al*. used native MS to confirm that encapsulin encapsulated just a single hexameric DyP and used native MS, atomic force microscopy, and multiscale computational modeling to demonstrate that cargo encapsulation has a substantial mechanical impact on the shell[239]. Encapsulin nanocompartments are gaining considerable interest from bioengineers because of their potential to allow the targeted delivery of drugs, proteins, and mRNAs to specific cells of interest[277,278]. For instance, Rurup *et al*. fused the DyP C-terminal end to the teal fluorescent protein (TFP), which by coexpression could be packaged within the encapsulin. High-resolution native MS was subsequently used to determine the cargo load and reveal the distribution of around 8-12 TFP proteins per encapsulin[279].

Another protein that naturally self-assembles is lumazine synthase (termed AaLS-WT), which forms a 60-mer assembly composed of 12 pentamers (T = 1, ~ 1 MDa)[280,281]. The group of Hilvert programmed electrostatic effects by incorporating designed mutations to tailor this protein, induce larger compartments, and optimize its assembly and cargo loading efficiency. They achieved a stepwise expansion of the *natural* protein shell, giving rise to thermostable ~ 3 and ~ 6 MDa assemblies containing 180 and 360 subunits, named AaLS-neg and AaLS-RR, respectively. High-resolution native (tandem) MS was used to probe the stoichiometry and topology of all of these variants[30,282]. The tandem CID-MS data for AaLS-neg are depicted in **Figure 15**. Increasing the collision energy revealed the consecutive elimination of up to 10 monomeric subunits. As these ejected monomers were highly charged, the resulting concomitant high-mass fragments became substantially less charged and even reached m/z = 70,000 Th[30]. Remarkably, as revealed by parallel analysis with cryo-EM, these expanded AaLS-neg and AaLS-RR particles retained tetrahedrally and icosahedrally symmetric structures constructed entirely from pentameric units[282], a relatively rare topology for virus-like particles. In a follow-up study, the T = 1, ~ 1 MDa AaLS-WT particle was also investigated by native top-down mass spectrometry. Its dissociation was induced by UVPD, with the loss of both monomeric subunits and sequence tags thereof, which could be used to identify the subunits[123].

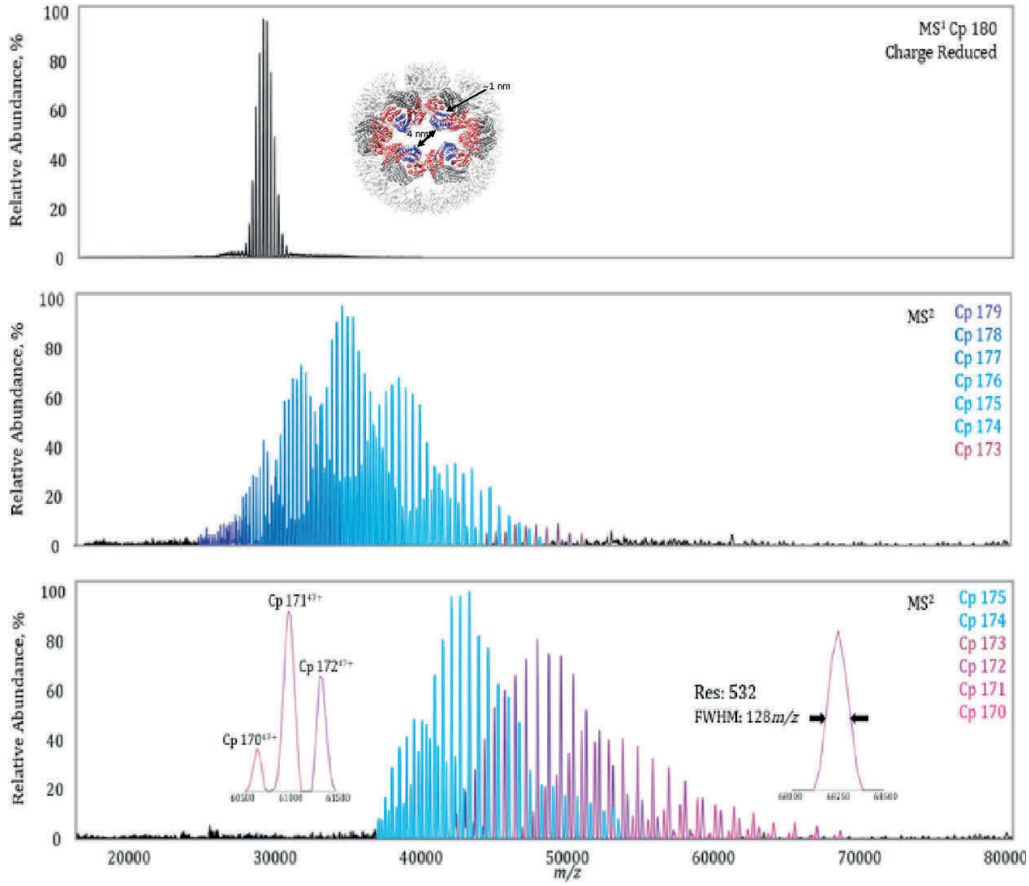


Figure 15 | High-resolution native tandem mass spectra of the designed charge-reduced AaLS-neg particles. (top) The charge-reduced intact Cp180 shows a well-resolved charge-state envelope centered at *m/z* ≈ 30,000 Th. (middle, bottom) The charge-state envelope was subjected to increasing HCD collision energies of 250 and 300 V. At maximal HCD the mass spectrum revealed the consecutive elimination of monomers from the Cp180-mer with fragment ions up to *m/z* = 70,000 Th. The obtained mass resolution enabled baseline resolution of these different dissociation products. At the highest *m/z*, the mass resolution was still greater than 500. The figure is adapted from ref [30]. Copyright 2017 Royal Society of Chemistry. The inset structure in the top panel is from ref [282]. CC BY 4.0.

In recent years, the groups of King[283] and Baker[284] have explored protein engineering and molecular self-assembly to generate de novo sophisticated self-assembling protein structures that even can be designed and used as nanoparticle vaccines against COVID-19[285]. For quality control, it is important to have the appropriate analytical tools to monitor their correct assembly and stability, whereby native MS can be a technique of choice. For instance, Wargacki *et al.*[286] combined high-resolution native MS with several complementary biophysical methods to monitor in detail the assembly of two of such computationally designed 120-subunit icosahedral protein cages that resemble VLPs. They used native MS to understand the assembly process and particularly its robustness toward perturbation. Also, for these designed protein cages, the assembly process was

highly cooperative and exclusively yielded complete 120-subunit complexes except in one nonstoichiometric regime for one of the materials. The latter could be uniquely monitored only by making use of the high-resolution native MS data. Related to this work, the group of Wysocki used high-resolution native MS to screen a library of designed building blocks for their propensity to form higher oligomers, defining their stoichiometry and topology[284,287,288].

Cumulatively, all these data reveal that high-resolution native MS is an excellent tool to characterize viruses and VLPs. Viruses (and bacterial encapsulins) not only are important as some of the most critical pathogens threatening (human) health but also are structurally very stable and therefore attractive biomolecular tools to be used in nanotechnology and bioengineering with applications in gene delivery and vaccine development. Their biochemical and biophysical properties are also fundamentally interesting, and native MS provides an ideal means to investigate their stability and assembly mechanisms.

Future developments in native MS will hopefully address the remaining challenges in structural virology, which are mainly posed by the mass heterogeneity of enveloped viruses, larger viruses containing a dozen different subunits (*e.g.*, adenovirus), or viruses containing extensively glycosylated spike proteins (*e.g.*, HIV and SARS-CoV-2). Such systems may readily be analyzed by CDMS[108,127,149,289], although to date it lacks the resolving power obtainable by conventional native MS as described here.

7.4. Protein Complexes Involved in Complement Activation

In this section, we focus on contributions that native MS has made to our understanding of some of the molecular pathways underlying immunity, in particular through the classical pathway of complement activation[290]. When initiated by target-bound antibodies, complement ensues as a cascade of proteolytic reactions that generate an array of bioactive compounds, including chemoattractants, opsonins, and anaphylatoxins. Together, these molecules provide signals to form ultimately membrane attack complexes (MACs) and to recruit immune cells for the clearance of pathogens and damaged host cells. The complement recognition complex C1 comprises six low-affinity binding arms that require antibody clustering to enable stronger multivalent binding, making it structurally compatible with naturally occurring antibody oligomers such as immunoglobulin M (IgM) pentamers and hexamers[291]. Although researchers have long known that IgGs, which occur mostly as monomers in serum, can also activate complement, the exact mechanism by which they do so remained elusive until just a few years ago.

In a landmark study, Diebolder *et al.*[292] showed that IgG activates complement by forming preferentially hexameric rings on antigenic surfaces. Based on the observation that some crystal structures of IgG show hexameric packing through Fc interactions, they used targeted mutations in the Fc-Fc interface that could enhance oligomerization and thus also complement activation. Native MS proved to be a key tool in this context, enabling them to confirm that a triple mutant, termed IgG1-RGY, readily forms non-covalent hexamers already in solution. Using cryo-EM, they next investigated the structure of target-bound

$(IgG1\text{-}RGY)_6$:C1 complexes, showing hexagonal rings of monovalent antigen-bound antibodies with C1 bound on top of the IgG platform. Encouraged by these observations, De Jong *et al.*[293] explored the use of hexamerization-enhanced IgGs as a therapeutic platform. Different constructs were screened using various methodologies, including native MS, to identify IgG mutants that remain in their monomeric inactive state in solution but display increased hexamerization and complement activation upon target binding.

Follow-up studies investigated the molecular mechanisms by which complement activation complexes are assembled. Using tandem MS on an Orbitrap Q Exactive EMR to study heterogeneous antibody-antigen complexes, Dyachenko *et al.*[119] showed that solution-phase IgG1-RGY hexamers can still bind up to 12 copies of their antigens. Interestingly, however, Wang *et al.*[294] revealed that antigen binding was not a requirement for C1 recruitment, as it binds readily with IgG hexamers in either the presence or absence of antigen. The role of antigen-binding was rather related to the activation of C1, likely by a conformational change induced through the IgG Fab arms. Finally, Wang *et al.* reconstituted a complex of C1 bound to an antigen-saturated IgG hexamer and used native MS to visualize the complete 2.1 MDa complement initiation complex (**Figure 16**). In a recent study by Strasser *et al.*[295], native MS was also used to corroborate high-speed atomic force microscopy (HS-AFM) studies into the mechanisms of IgG hexamerization on antigenic surfaces. Two main assembly pathways were identified, the more important of which was a "vertical pathway" in which target-bound IgG recruits additional IgG monomers from solution, followed by monovalent antigen binding. Although intermediate oligomers were also observed, at least five binding sites were required for C1 binding and activation. These reports have improved our mechanistic view into the initial steps of the complement cascade and highlight the complementary and rather unique role of native MS in hybrid structural biology.

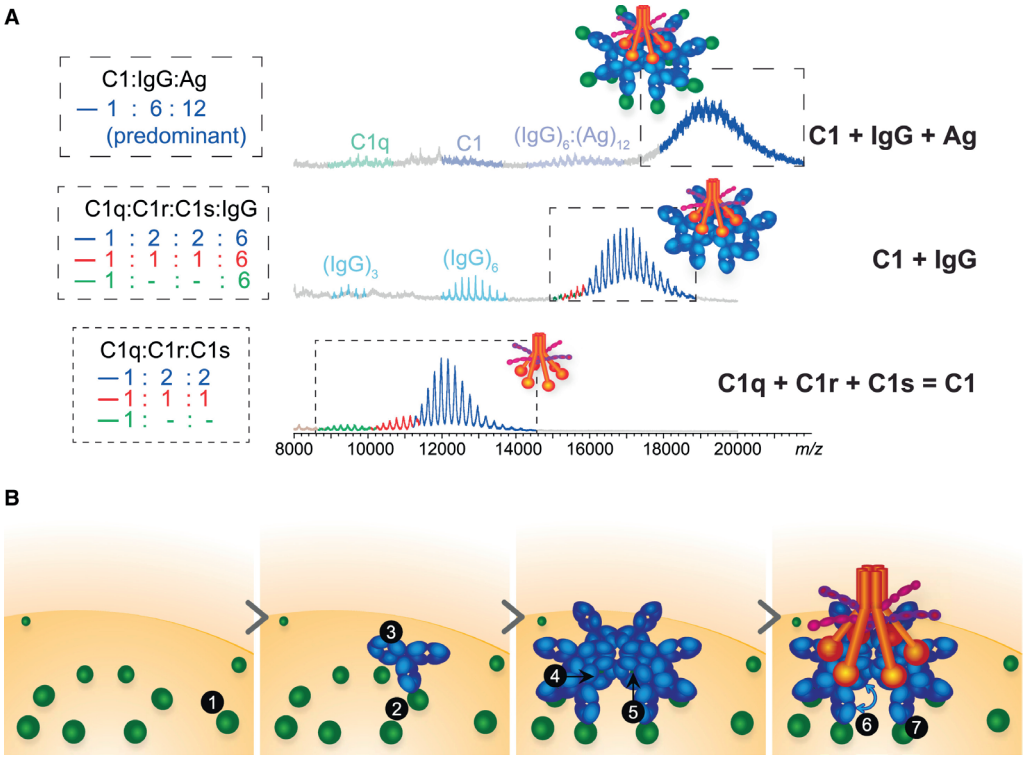


Figure 16 | Native-MS-assisted analysis of the initial steps in complement activation: stepwise reconstitution of the 2.1 MDa C1:IgG:Ag complexes, comprising 40 protein subunits, monitored by native MS. (A) Native MS analysis of reconstituted C1, C1:IgG, and C1:IgG:Ag complexes. The signals shown in dashed boxes are color-coded according to the stoichiometries of C1 assembly, as specified to the left of the corresponding spectra. The bottom spectrum demonstrates the successful reconstitution of C1, containing C1q, two copies of C1r, and two copies of C1s. The middle and top spectra show that the latter assembly successfully binds to $(IgG)_6$ and subsequently to 12 copies of the antigen (Ag). **(B)** Model summarizing molecular determinants contributing to IgG-mediated activation of the classical complement pathway. The assembly of the complement initiation complex is proposed to be dependent on (1) the availability of specific antigens recognized by IgG antibodies, (2) Ag compatibility with clustering at the cell surface or in solution, (3) the Fc-domain conformation, (4) the high-avidity binding site for hexavalent C1q, (5) the composition of the IgG glycan chains, (6) the presence of Fab arms, and (7) Ag binding by Fab arms. Reproduced with permission from ref [294]. Copyright 2016 Elsevier Inc.

7.5. Protein Assemblies Involved in Light Harvesting

The conversion of light into energy is one of the most fundamental biological processes on earth, and through billions of years of evolution it has been optimized to great efficiency. The most efficient light-to-energy conversion occurs within the outer membranes of some microbes and plants and is mostly mediated by very large photosynthetic protein assemblies. These photosynthetic protein complexes are mediated by well-structured and protein-mediated networks of light-absorbing pigment molecules. Studying their structures remains a key prerequisite to understanding better the process of light, which is of tremendous fundamental importance and is crucial for the development of novel artificial light-harvesting machineries[296]. The main advantage of utilizing native MS in the

analysis of protein assemblies involved in photosynthesis is its ability to preserve non-covalent interactions, which mediate not only interactions of protein subunits within a photosynthetic complex but also the covalent and non-covalent binding of various cofactors and multiple pigments (*e.g.*, chlorophylls). Even when these cofactors are bound covalently, as in the case of phycobiliproteins, the bonds are typically weaker than the covalent bonds in the proteins, making soft ionization conditions very important in the analysis of photosynthetic protein assemblies. Because of the varying occupancy of cofactors or pigments bound per single photosynthetic complex, high resolution is often required to discern between complex variants with native MS. Over the past decade, native MS alone and as a complementary tool was successfully employed to analyze various photosynthetic protein complexes to study pigment occupancies, oligomeric states, and stabilities and even to define photoprotective and light-harvesting mechanisms.

A large amount of the earliest work that leveraged the structural analysis of photosynthetic proteins with native MS was performed by the Blankenship group[29,297-304]. Their first attempt at using native MS to characterize the well-studied Fenna-Matthews-Olson (FMO) complex, the first pigment-protein complex with a solved crystal structure[305,306], complemented available structural data by determining the exact number of pigment molecules bound within the complex. In their work, Wen *et al*. took advantage of native MS to provide a snapshot of complex heterogeneity and confirmed that one of the pigments missing in hitherto available structures demonstrated partial occupancy due to surface exposure[304]. In a subsequent study, Zhang *et al*. employed ECD on an FT-ICR mass spectrometer under native conditions to determine flexible regions within the FMO protein subunits[29]. Further work was done on the orange carotenoid protein (OCP), which protects cyanobacteria from photodamage by close interaction with the light-harvesting phycobilisome (PBS)[303]. The sensitivity of native MS allowed them to detect the protein in a complex with its pigments and investigate the light-dependent oligomerization of OCP, which is essential for its photoprotective function. In these studies, some of the inherent limitations of native MS in providing high-resolution structural information were acknowledged. However, they also demonstrated the prospects of combining multiple mass spectrometric modalities (*e.g.*, cross-linking MS and native MS) to analyze these complex multicomponent light-harvesting assemblies. In the follow-up studies, native MS was further used in combination with collisional unfolding in the analysis of OCP, revealing structural features of its functional domains and compensating for the absence of high-resolution structures[301].

The oligomeric state, pigment content, and complex stabilities were also determined for other photoprotective proteins, like peridinin-chlorophyll-protein[302] and the fluorescent recovery protein (FRP)[299]. By combining native MS with cross-linking MS and site-specific mutations, Lu *et al*. proposed a novel mechanism of function for dimeric FRP, which participates in photoquenching by dissociating OCP from the PBS[300]. When native MS matured for the analysis of membrane-embedded proteins, it was applied to characterize membrane photosynthetic reaction center (RC) complexes[297,298]. Although membrane complexes are electrosprayed in the presence of solubilizing detergents and require additional collisional activation to be released from the micelle, RC complexes could still

be detected with multiple non-covalent cofactors attached, further demonstrating the ability of native MS to retain even labile non-covalent interactions (**Figure 17**).

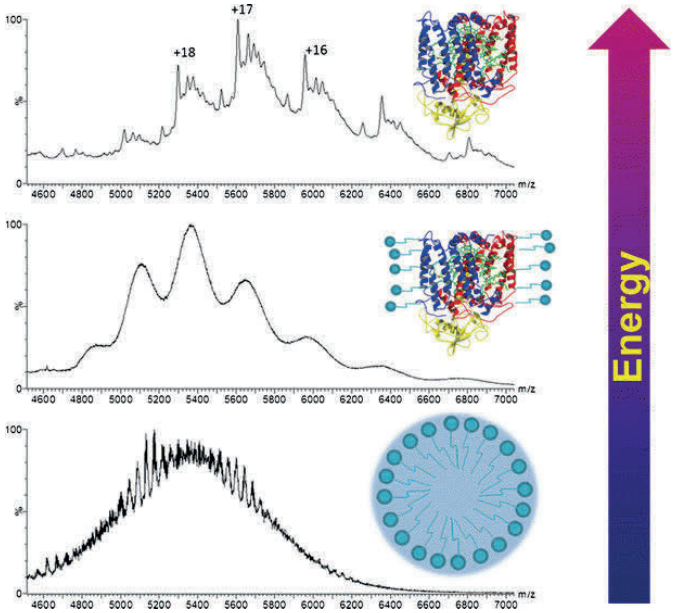


Figure 17 | Native MS analysis of a membrane-embedded photosynthetic reaction center: release of the photosynthetic reaction center complex from detergent micelles by collisional activation. At low collisional energy (bottom spectrum), the broad peak represents ions from the micelles. The detergent micelles and extra bound detergent can be removed by increasing the collision energy at the trap region of a Synapt G2 mass spectrometer. Reproduced from ref [298]. Copyright 2017 American Chemical Society.

More recently, with increasing interest in designing even more efficient light-harvesting proteins, various components of the main light-harvesting machinery in cyanobacteria and red algae, called phycobiliproteins, have been investigated. Although in the last years cryo-EM – following the so-called "resolution revolution" – has yielded exciting and detailed structural studies of entire PBS systems from different origins[307,308], native MS has provided a unique angle, exposing heterogeneities observed in these complexes and exploring complexation dynamics under various conditions. Leney *et al*. investigated the stabilities and fluorescence of heptameric B-phycoerythrin (BPE), the most abundant PBS subcomplex in red algae. Since phycoerythrins purified from red algae represent some of the brightest fluorophores on the planet and can be commercially obtained in large quantities, it is also important to determine their stabilities upon environmental perturbations. By taking snapshots of BPE complexes with native MS under different pH conditions, Leney *et al*. tracked the complex dissociation and assembly pathways, determining that the complete heptameric form is essential for its distinctive photochemical properties[309]. In the last few years, several other studies employing native MS as the core technique for analyzing phycobiliproteins have been reported[123,310-313].

Although the use of native MS by itself has provided great insight into the biology and light-harvesting mechanisms of several photosynthetic assemblies, the technique is more and more complemented by other MS modalities, especially top-down and bottom-up MS and IM tandem MS. For instance, to explain the anomalous heterogeneity of BPE observed by high-resolution native MS that could not be explained by the expected pigment content and oligomeric state, Tamara *et al*. opted for top-down analysis of BPE proteoforms to chart the masses of all of the assembly variants, ultimately assigning the observed heterogeneity to the numerous co-occurring forms of the core g subunit[311]. Furthermore, they discerned the isobaric yet distinctive pigment molecules via bottom-up HCD-MS/MS analysis and subsequently assessed each pigment's total content in the native complexes. A high degree of sequence information and partial sequence positioning was achieved for entire BPE assemblies with native top-down MS via UVPD[123]. In another example, native MS combined with IM-MS allowed the stabilizing role of the critical subunits in the phycobilisome proteins[310]. In a study on UVR8 plant photoreceptors, Camacho *et al*.[314] explored the functional role of the oligomeric state in the photoprotection mechanism with native IM-MS and collisional unfolding (**Figure 18**). Such strategies, supplemented by improvements in desolvation, transmission, and resolution of instruments for native MS, can open new avenues for analyzing other photosynthetic protein complexes of ever-increasing size and complexity, such as Photosystem II supercomplexes[315].

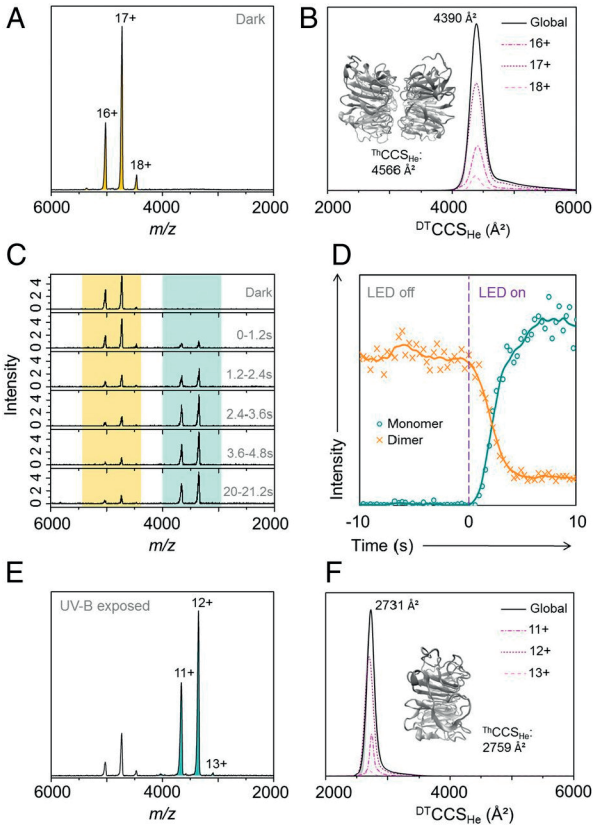


Figure 18 | Native MS and IM-MS of the UVR8 photoreceptor core domain (UVR8$^{12\text{-}381}$). (A) Native mass spectrum of the UVR8$^{12\text{-}381}$ dimer (pale-orange peaks) in the absence of UV-B light. **(B)** Monodisperse drift tube collision cross section in helium ($^{DT}CCS_{He}$) of the UVR8 dimer. The inset shows the energy-minimized structure of the UVR8 dimer and associated theoretical CCS ($^{Th}CCS_{He}$). **(C)** Mass spectra of UVR8$^{12\text{-}381}$ as a function of illumination time (280 nm, 25 mW, 350 mA). Mass spectra were combined over a period of 1.2 s after the light-emitting diode (LED) was switched on at t_0. Dimer signal, pale orange; monomer signals, pale green. **(D)** Ion chromatograms extracted for the UVR8$^{12\text{-}381}$ dimer (orange) and monomer (green) as functions of acquisition time. Data to the left of the purple dashed line are from when the ion source tip was in the dark, and the data to the right are from when the tip was illuminated. **(E)** Native mass spectrum of UVR8$^{12\text{-}381}$ following illumination in the source tip with the 280 nm LED for 10 s (to ensure maximum conversion). The spectrum is now dominated by the UVR8$^{12\text{-}381}$ monomer (pale-green peaks). **(F)** Monodisperse $^{DT}CCS_{He}$ of the UVR8$^{12\text{-}381}$ monomer. The inset shows the energy-minimized structure of the UVR8$^{12\text{-}381}$ monomer and associated $^{Th}CCS_{He}$. From ref [314]. CC BY 4.0.

7.6. Membrane-Embedded Protein Complexes

In recent years, native MS has emerged as a versatile and highly sensitive technique to study membrane proteins and their interactions, almost becoming a separate discipline. For a long time, membrane proteins were thought to be intractable by electrospray native MS because of their intrinsic hydrophobic nature, which makes them incompatible with the aqueous solutions used in standard electrospray. Now that the solutions to circumvent this incompatibility have been developed (see below), the application of native MS to the investigation of membrane proteins has been greatly expanding. Several excellent and

recent reviews extensively cover the development in this field[20,195,196,316,317], showing the specific set of challenges mass spectrometrists face when analyzing membrane proteins and their complexes. Membrane proteins and their interactions are of major biological importance, as they contribute roughly 30% to the human proteome and are targeted by as many as 50% of all currently used drugs. Although they have key functions in signaling, transport, metabolism, and the respiratory chain, this class of proteins remains understudied in vitro compared with their soluble counterparts. This is the case because membrane proteins require membranes or mimetics thereof to remain in a functional native-like state, making them not only harder to produce recombinantly but also more challenging to tackle analytically using traditional biophysical techniques. For native MS analysis, much effort has gone into the development of MS-compatible membrane mimetics to transfer membrane proteins into the gas phase as well as instrumental modifications that enable more extensive ion activation to strip away resulting unwanted adducts. In keeping with the scope of this review, we will briefly highlight some of the most recent advances.

Partly driven by skepticism about the biological relevance of using detergent micelles, advances in nanodisc technology have enabled researchers to solubilize membrane proteins in a more native-like environment. Nanodiscs are lipid bilayers of which the hydrophobic edge is covered by two copies of amphipathic membrane scaffolding protein (MSP), providing a homogeneous and monodisperse model system. However, initial native MS studies required exceptionally high levels of collisional activation to release the "naked" membrane proteins[318], while the stability of the nanodiscs was too low to measure them intact[319,320]. To overcome these challenges, Keener *et al*.[321] explored the use of chemical charge modulation additives to tune the gas-phase stability of the used nanodiscs. Using macromolecular mass defect analysis to assign stoichiometries, they determined the oligomeric states of ammonium transport protein (AmtB) and AqpZ in intact nanodiscs without potential distortions introduced by collisional activation. In two follow-up studies, Walker *et al*.[322,323] utilized the same method to characterize the interactions of antimicrobial peptides (AMPs) with nanodiscs modeling mammalian or bacterial membranes. AMPs were titrated into preformed nanodiscs to promote insertion into the membrane, followed by direct measurement via native MS. Unique stoichiometries were observed for each AMP depending on the lipid type used and the AMP concentration. Moreover, collisional activation experiments could distinguish between the formation of specific membrane-destabilizing complexes and nonspecific binding, demonstrating the potential of native MS to provide key insights into AMP mechanisms.

In two recent papers, Chorev *et al*.[324,325] demonstrated the mass analysis of protein complexes from lipid vesicles derived directly from native cellular membranes. By sonication of large membrane fragments isolated from cells, they were able to produce small vesicles of similar lipid and protein composition without the need for recombinant overexpression or further purification. Native mass spectra were obtained using a modified Orbitrap UHMR, although the heterogeneity of the proteins and bound lipids presented a significant challenge for mass assignment. They therefore employed small-molecule MS, lipidomics, proteomics, and molecular dynamics simulations to gain more

confidence in their mass assignments. This combination of methods enabled them to report on multiple protein assemblies present in the *E. coli* membrane (**Figure 19**) and those present in bovine mitochondrial membranes, some of which had not been previously described. Although some skepticism remained about their analysis and in particular their mass assignments[326,327], this advancement shows that native MS has the potential to provide pivotal data when it comes to membrane proteins in a very native-like context.

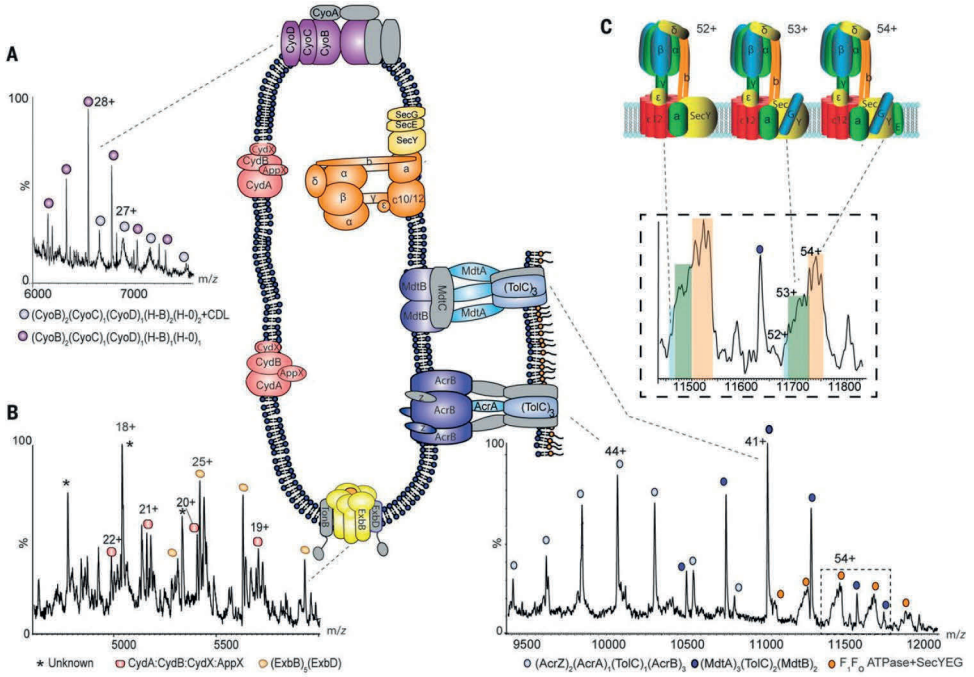


Figure 19 | Native MS of membrane-embedded protein complexes analyzed directly from their native environment. Regions of the mass spectrum recorded for inner membranes from *E. coli* yield cytochromes, the Ton complex multidrug transporters, and the intact ATP synthase in complex with the SecYEG translocon. **(A, B)** Expanded regions of the spectrum assigned to cytochrome bo_3 and cytochrome *bd* oxidase, showing peak splitting due to binding of quinol and heme groups. The pentameric ExbB complex (with one copy of ExbD in the center of the pore) that forms part of the TonB complex is also observed (yellow). **(C)** High-*m/z* region of the mass spectrum assigned to the multidrug efflux pumps AcrAB and MdtAB and the intact ATP synthase. Expansion of peaks assigned to the ATPase reveals binding of the SecY (blue), SecYG (green), and SecYEG (orange) charge states 52+, 53+, and 54+. Complexes observed in mass spectra are shown schematically, with subunits that have dissociated shown in gray. Reproduced with permission from ref [324]. Copyright 2018 Chorev *et al*.

8. Proteoform Profiling by High-Resolution Native MS

In recent years, high-resolution native MS has also emerged as a powerful technology for proteoform profiling, exposing the microheterogeneity induced by protein glycosylation and additional PTMs. The rapidly expanding field of mass-spectrometry-based glycoproteomics relied until recently primarily on two levels of analysis, targeting either released *N*-glycans or analyzing proteolytically formed and enriched glycopeptides[328-330]. Although very powerful, these approaches also have limitations, mainly caused by the enormous structural microheterogeneity that most glycoproteins exhibit[331]. Therefore, there is a dire need for additional more protein-centric approaches.

In the analysis of intact proteins, native MS has some well-defined advantages compared with denaturing (conventional) MS also when it comes to proteoform profiling. Native MS provides an improved view of sample heterogeneity through the characteristically increased spacing between adjacent charge states and can even boost sensitivity by minimizing the number of charge states over which the signal is distributed over[53]. Moreover, less charged and more globular molecules measured under native MS conditions are less prone to decay during longer signal acquisition times, which for FT MS instruments are essential to achieve high mass resolution[40]. These features make high-resolution native MS an excellent method for revealing the glycoform profiles of glycoproteins qualitatively and quantitively[332].

In this section, we begin by describing how native MS first contributed to the analysis of intact glycoproteins. We then discuss the emerging role of native MS in characterizing microheterogeneities in therapeutic antibodies and other biopharmaceuticals. Next, we discuss advancements in MS-based profiling of plasma glycoproteins and finally review work on proteins harboring other PTMs, such as phosphorylation.

8.1. Native MS of Intact Glycoproteins

The enormous structural heterogeneity of intact glycoproteins is the major bottleneck in their effective characterization also when mass spectrometry is used. Early studies analyzing intact glycoproteins, such as ovalbumin, by denaturing mass spectrometry suffered from poorly resolved ion signals, hampering accurate mass assignments and therefore in-depth glycoprofile analysis[333]. Not surprisingly, as a widely available model system for analytical technology development, ovalbumin was also one of the first glycoproteins analyzed by high-resolution native MS shortly after the introduction of the Orbitrap EMR. Using high-resolution native MS, Yang *et al*.[334] revealed that ovalbumin exhibits a diverse glycoproteoform profile even though it harbors only a single *N*-glycosylation site. This is the case because two additional phosphorylation sites present on ovalbumin create quite a few additional proteoforms. Therefore, Yang *et al*. also used native MS to analyze ovalbumin treated with a phosphatase and/or an endoglycosidase to selectively cleave off the phosphate and glycan moieties, providing simplified proteoform profiles that facilitated the annotation of around 60 distinguishable proteoforms. The detected repertoire of ovalbumin proteoforms was recently extended even further by Füssl *et al*.[335], who subjected intact ovalbumin to online anion-exchange chromatography

using a pH gradient elution coupled to mass spectrometric detection under native conditions and reported more than 150 different proteoforms. Already these data highlight certain advantages of analyzing intact glycoproteins compared with peptide-centric approaches, as the latter are hampered by simultaneous detection of unmodified and glycosylated peptides as well as peptides decorated with other PTMs (*e.g.*, phosphorylation).

Another glycoprotein that has often been tackled by new analytical approaches and therefore evidently also by high-resolution native MS is bovine fetuin[336,337]. The glycosylation profile of bovine fetuin is quite well understood, as it has been exposed by a plethora of glycan- and glycopeptide-centric studies. Intriguingly, Lin *et al*.[338] used high-resolution native MS to directly compare the glycoproteome profiles of bovine serum fetuin, human serum fetuin, and recombinant human fetuin expressed in HEK-293 cells. The native MS data for these three similar proteins displayed considerable differences in their proteoform profiles due to differences in protein maturation, phosphorylation, and *N*- and *O*-glycosylation. Although the two *N*-glycosylation sites, the *O*-glycosylation site, and the phosphorylation site are conserved from bovine to human, the stoichiometry of the modifications and the specific glycan structures they harbor were found to be quite distinct. By comparison of serum and recombinant human fetuin, it was observed that fetuin from serum exhibited a much simpler proteoform profile, indicating that the recombinant protein is not ideally engineered to mimic human serum fetuin, as is likely true for many more recombinant glycoproteins currently in use in various biochemical studies. As these glycoproteoform profiles may have functional consequences, such studies should therefore be evaluated carefully.

Compared with ovalbumin and even fetuin, nature has created many glycoproteins that are much more complicated. Notably, several viruses contain spike proteins that are heavily glycosylated in the form of a glycan shield, which directly impacts antibody binding and interactions with the host cell [339,340]. Struwe *et al*.[341] used high-resolution native MS to probe the *N*-glycan site occupancy of the HIV-1 gp120 glycoprotein. Using gp120 expressed in HEK293T cells in the presence of the α-mannosidase inhibitor kifunensine to reduce the heterogeneity in glycosylation, they were able to determine the occupancy of several glycosylation sites. However, it remains a future challenge to perform such analyses on wild-type gp120, ideally extracted from endogenous viral particles. As protein glycosylation has such an impact on the virus-host interaction, studies that target viral glycoproteins will likely become more prominent in the future.

8.2. Therapeutic Antibodies

Over the past decades, biopharmaceutical therapies have rapidly emerged to dramatically change the treatment of diseases, with monoclonal antibodies (mAbs) currently dominating biopharmaceutical approvals and sales[342]. As complex biomolecules, biopharmaceuticals are amenable to a variety of PTMs, including complex glycosylation, glutamine cyclization, and C-terminal lysine clipping. Many of these modifications are considered critical quality attributes (CQAs) because they can influence the safety and efficacy profile of the drug. Notably, protein *N*-glycosylation is known to modulate the

function of biopharmaceuticals in a variety of ways. For instance, antibody-dependent cell-mediated cytotoxicity is significantly increased when the *N*-glycans in the Fc domain of IgG molecules lack a fucose residue[343]. Additionally, for therapeutic antibodies, galactosylation and sialylation play a role in complement-dependent cytotoxicity[344] and anti-inflammatory activity[345], respectively. As protein glycosylation is very much dependent on the particular host cell and the specific conditions employed for production, its pattern provides a means to distinguish the original patented drug from biosimilars[346-348].

8.2.1. High-Resolution Native MS Provides a Snapshot of Antibody Microheterogeneity

High-resolution native MS has become an integral technology in the toolbox for exposing the microheterogeneity of antibody products[349-351]. With the first demonstrations of the Orbitrap platform for native MS, Rose *et al*.[27] and Rosati *et al*.[31] immediately noted that the ToF instruments used up until that point could be outperformed substantially when it comes to resolving antibody microheterogeneity. From a single mass measurement, different glycoforms of intact ~ 150 kDa IgG1 antibodies could be baseline-resolved with molecular weights differing by 162 Da, representing distinct numbers of hexose units incorporated into *N*-linked glycan structures. Mass differences down to 25 Da could be sufficiently resolved, opening the door to the identification of smaller PTMs and amino acid mutations.

The methods by which native MS can provide detailed analysis of microheterogeneity in antibodies were extensively described in two papers by Rosati *et al*.[117,352]. One of the key advantages of using native MS for this purpose is that it allows for increased charge spacing compared with denaturing MS, extending the level of heterogeneity and sample complexity that can be studied. The first step in analyzing monoclonal antibodies by native MS is generally to determine their backbone mass profile after enzymatic removal of *N*-linked glycans using PNGase F. Alternatively, neuraminidase and β1,4-galactosidase may be used to selectively remove sialic acid and free galactose moieties, reducing the complexity of glycans and therefore the resulting native MS spectra. This step is essential to discern double fucosylation from sialylation since these modifications amount to similar mass shifts (292.3 and 291.3 Da, respectively). The total contribution of glycosylation to the molecular weight is readily assessable by comparison of the mass spectra of antibodies before and after complete deglycosylation. Rosati *et al*.[117] showed that their workflow enables both qualitative and quantitative analysis of composite glycosylation profiles and other sources of microheterogeneity, in some cases identifying over 30 proteoforms of a single antibody.

The simplicity of sample preparation and speed of analysis has made high-resolution native MS a method of choice for initial screening of antibody batches during development and production. Using protein chemistry for in vitro glycoengineering of trastuzumab, Parsons *et al*.[353] utilized high-resolution native MS to study the reaction products. After selective endoglycosidase-catalyzed glycosylation in the presence of an oxazoline donor,

native MS revealed not only the expected product but also the occurrence of unwanted nonselective glycation reactions. Native MS enabled them to direct the optimization of their strategy, producing highly pure trastuzumab samples with specific natural or non-natural glycans.

In another antibody engineering approach, van der Schoot *et al*.[354] recently presented a CRISPR/HDR platform to rapidly engineer immunoglobulin constant domains and to produce recombinant hybridomas that excrete these antibodies in the preferred format, species, and isotype. High-resolution native MS enabled them to rapidly characterize antibodies with complex microheterogeneities, confirming for example that an Fc-silent N297A mutant IgG2 indeed lacked *N*-glycosylation, diminishing its ability to recruit effector functions through FcγR binding.

In another study, Thompson *et al*.[355] characterized oligoclonic mixtures of antibodies produced in a single production platform, demonstrating that up to 15 unique IgGs could be identified and quantified unambiguously and simultaneously. Native MS can even be used to characterize overproduced proteins directly from crude samples from expression systems, as was shown by Vimer *et al*.[356]. By comparison of the growth media of HEK293F cells for nonexpressing and expressing conditions, peaks corresponding to a secreted antibody could be assigned, thereby exposing its glycosylation profile. Combined, these studies highlight the promising role of native MS as a screening tool in optimizing expression conditions and studying batch-to-batch variations in therapeutic antibodies.

When native MS is being used for glycoform "fingerprinting" of antibodies, care should be taken regarding the occurrence of hard-to-discern isobaric variants. The most common of such modifications is glycation, which can reduce the efficacy and can even render the antibody immunogenic. Non-enzymatic attachment of a hexose to a backbone residue results in a mass shift identical to that for a mannose or antennal galactose extension of an *N*-glycan, making it indiscernible on the intact-protein level. Recently, Esser-Skala *et al*.[357,358] presented a method to assign possible PTM compositions and to eliminate hexosylation bias in the analysis of *N*-glycosylation patterns by native MS. When measuring the antibody in its *N*-deglycosylated form, satellite peaks at intervals of 162 Da indicated the presence of nonspecifically glycated isoforms. *N*-Glycoform abundances in the native mass spectrum of the original product could then be computationally corrected by construction of a glycation graph that gathers all possible abundance transfers between proteoforms. Esser-Skala *et al*. validated their approach on various antibody samples, demonstrating for example that although the apparent *N*-glycan abundances in native mass spectra of two batches of denosumab were highly similar, different levels of glycation led to substantial differences in their corrected *N*-glycan profiles (**Figure 20**).

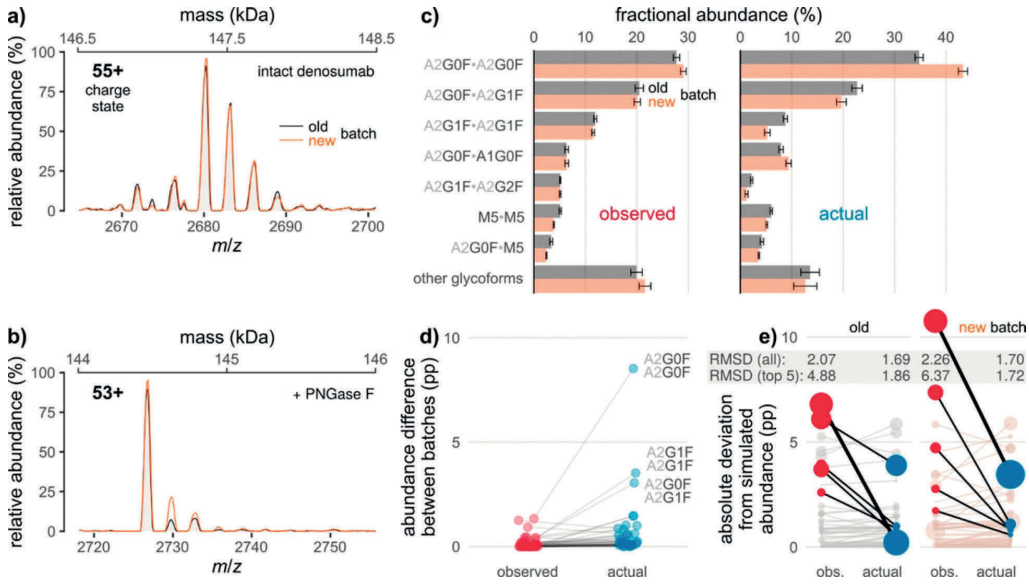


Figure 20 | Eliminating hexosylation bias in glycoproteoform profiling by high-resolution native MS. Glycation obscures differences between the glycoform profiles of two batches of Prolia (old vs new). **(a, b)** Raw mass spectra of the **(a)** intact and **(b)** de-*N*-glycosylated mAb, corresponding to 2 kDa sections of the respective zero-charge spectra (the secondary *x* axes indicate the respective masses). **(c)** Fractional glycoform abundances before and after correction for the effects of glycation. **(d)** Interbatch differences of glycoform abundances as derived from (**c**). Lines connect points denoting identical glycoforms. The three most common glycoforms are labeled. pp, percentage points. **(e)** Absolute deviations of observed and actual glycoform abundances from simulated abundances based on released *N*-glycan data. In each batch, those five glycoforms for which correction leads to the largest decrease in deviation are highlighted (thick line: most pronounced decrease). Point areas are proportional to observed/actual glycoform abundance. Error bars represent (propagated) 95% confidence intervals from five technical replicates. RMSD, root-mean-square deviation. From ref [357]. CC BY 4.0.

8.2.2. Hybrid Approaches Localize and Detail the Sources of Microheterogeneity

While native MS can rapidly provide a snapshot of the ensemble of complex proteoforms, especially when combined with bias correction algorithms, more information is needed to gain an in-depth understanding of glycan structures and PTM localization. Hybrid mass spectrometry approaches therefore combine native MS with additional information at the level of released glycans or glycopeptides. Combining native MS with released glycan analysis by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD), Rosati *et al.*[117] studied the effects of CH3 domain mutations on N297 glycosylation in the distant CH2 domain of IgG4 half molecules (IgG4Δhinge). In particular, mutations of Y407 by charged residues or glutamine stood out, as these mutants displayed a dramatic increase in galactosylation, chain branching, and sialylation. Translating the Y407E mutation to IgG1 resulted in similar changes in the glycosylation profile of this subclass, with HDX experiments revealing that structural changes may be responsible for this effect. In another hybrid approach, Yang *et al.*[332] profiled released glycans by nanoLC-Chip-Q-ToF MS, followed by mass and retention-time

matching to a previously reported human serum *N*-glycan structural library. An in silico-constructed zero-charge mass spectrum of the antibody was then compared to the real native MS data qualitatively and quantitatively. Although the approaches showed good agreement for the glycosylation profiles of IgG4Δhinge Y407 mutants, their combination proved to provide the most reliable and thorough characterization of mAbs. Other types of modifications could also be revealed. By measuring N-deglycosylated mAbs, Yang *et al*. showed that trastuzumab has a relatively monodisperse backbone, whereas infliximab displays variable degrees of C-terminal lysine clipping and bevacizumab is extensively glycated.

Hybrid approaches incorporating glycopeptide analysis can complement native MS analysis by localizing glycans and other PTMs, determining site occupancy, and revealing the compositional makeup of glycans. Recently, Brücher *et al*.[359] applied such an approach to investigate potentially distinctive glycosylation profile of antibodies produced in malignant tissues via gene delivery. To this end, they expressed two therapeutic mAbs with different modes of action in various cancer cell lines, both responsive and nonresponsive to the mAb produced, and then compared the glycosylation profiles to that of the same mAb expressed in standard HEK293 and CHO-S producer cells. They found that both the producer tissue and the antibody isotype influence sialylation and fucosylation and therefore the efficacy of the antibody, enabling them to identify optimal cell types according to the desired mode of action. Furthermore, they found that relatively high amounts of non-glycosylated antibodies were produced by cell lines responsive to the antibody, potentially decreasing the therapeutic efficacy of antibodies that function through antibody-dependent cell-mediated cytotoxicity (ADCC) because they depend on glycosylation for FcγR engagement. Native MS thus revealed substantial differences in microheterogeneity between producer cells that affect antibody efficacy, highlighting the importance of in vivo glycoengineering for the development of improved anticancer antibodies that are produced at the target tissue through gene delivery.

8.2.3. Online Separation Enables More Extensive Characterization of mAb Charge Variants

Because many PTMs affect the surface charge distribution of proteins, antibodies frequently exhibit a variety of charge variants that differ from one another in terms of efficacy and half-life. Therefore, antibody proteoforms can often be separated by charge-sensitive methods, which improves the sensitivity and the ability to resolve (nearly) isobaric proteoforms[350,360]. Although the elution buffers traditionally used in these approaches are incompatible with MS, limiting them to offline fractionation, recent efforts exploring volatile alternatives have enabled direct hyphenation to high-resolution native MS. By coupling pH gradient-based weak cation exchange chromatography (WCX) directly to an Orbitrap instrument, Bailey *et al*.[361] demonstrated effective native separation of antibody charge variants based on minimal differences in their isoelectric points. This enabled them to characterize heterogeneities that result in minimal mass shifts (*e.g.*, deamidation, ΔM = 1 Da) and even isobaric proteoforms (*e.g.*, due to aspartate isomerization), in both a qualitative and quantitative manner. In a follow-up study by

Phung *et al*.[362], a similar approach was used to study mispairing in bispecific IgGs produced by a single-cell host, providing a unique solution to resolve isobaric species. An alternative approach, hyphenation of strong cation exchange chromatography (SCX) to native MS, was presented by Füssl *et al*.[363], who characterized charge variants of several therapeutic antibodies using tailored gradient slopes for each mAb. Similarly, Yan *et al*.[364] extensively studied microheterogeneities in the NISTmAb. Notably, due to the high sensitivity of the approach, a non-consensus Fab glycosylated isoform could be identified at < 0.1 % abundance. In another approach, Ma *et al*.[365] complemented native SCX-MS with middle-up proteomics for subdomain analysis of several commercially available mAbs, exposing an array of microheterogeneities and changes in their abundance upon stressing of the sample. Although ion-exchange chromatography shows particular promise for exposing antibody microheterogeneity, alternative LC separation methods are also being explored for hyphenation to native MS. This includes in particular hydrophobic interaction chromatography (HIC)[366] and SEC[367,368].

Because of its ability to achieve fast separation paired with low sample consumption, capillary zone electrophoresis has emerged as an attractive MS-compatible alternative to LC separation[369]. Although CZE was originally used under denaturing conditions, two papers by Belov *et al*.[370,371] demonstrated one of the first implementations of CZE hyphenated to native MS. Using a neutral polyacrylamide-coated capillary coupled to an Orbitrap EMR, they analyzed various native proteins and protein complexes, including antibodies. By combining this approach with middle-down and intact CZE-MS analysis under denaturing conditions, they were able to identify and quantify multiple antibody charge variants and confirm PTM site localizations. Using a commercially available microfluidic CE device, Füssl *et al*.[372] and Carillo *et al*.[373] recently characterized charge variants of various therapeutic antibodies by native MS. Notably, they compared the performance of the CE platform to cation-exchange chromatography (CEX) in the analysis of cetuximab, an exceptionally heterogeneous IgG1 that harbors additional glycosylation sites in the Fab region. Using glycopeptide mapping to facilitate peak annotation, they identified and quantified over 200 proteoforms of cetuximab by CE-MS, twice as many as by CEX-MS (**Figure 21**). Combined with inherently low sample consumption, this makes CE-MS a promising tool for the analysis of therapeutic antibodies, also in the early stages of drug development when less material is available.

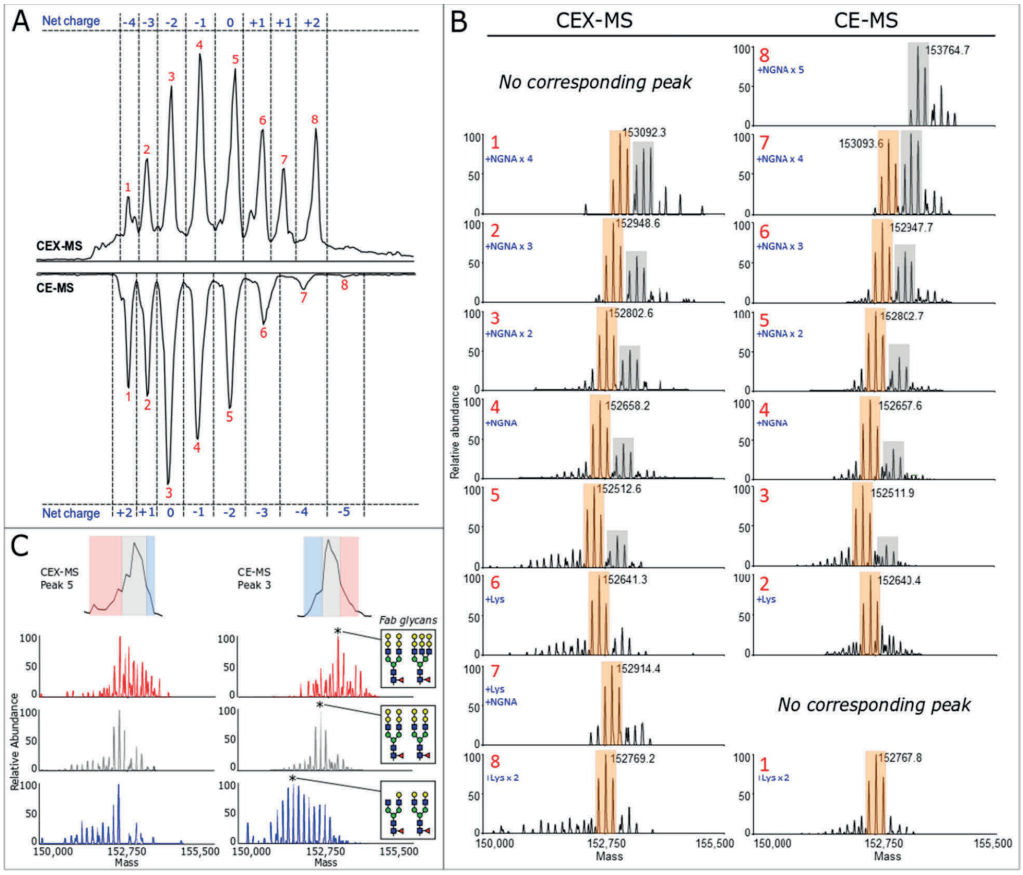


Figure 21 | Comparison between cation-exchange chromatography MS (CEX-MS) and capillary electrophoresis MS (CE-MS) in the analysis of cetuximab charge variants. (A) Cetuximab chromatogram and electropherogram, with peaks annotated in red and the corresponding net charges in blue. **(B)** Deconvoluted native mass spectra corresponding to the peaks in (A). Spectra from the two separation approaches containing the same charge variant species are aligned horizontally. Modifications that cause differences in the net charge are indicated in blue, whereas only the most probable combination of modifications is given. The most abundant species corresponding to diantennary Fab glycan pairs are shown in orange, whereas the most abundant forms carrying triantennary Fab glycans are labeled in gray. **(C)** Deconvoluted spectra of the peak front, center, and tail of peaks 5 and 3 of the CEX and CE separations, respectively. Associated spectra are aligned horizontally and are shown in the same color. Reproduced from ref [372]. Copyright 2020 American Chemical Society.

8.3. Antibody-Drug Conjugates

Next to the more conventional therapeutic antibodies described in the previous section, ADCs have recently become another important class of cancer biotherapeutics[374]. Some ADCs approved for clinical use (mostly in cancer) include Adcetris, Kadcyla, and Mylotarg. ADCs are typically composed of a monoclonal antibody to which a potent cytotoxic drug is attached via a cleavable linker. The specificity of the mAb component of the ADC to cancer antigens expressed on the cell surface ensures that the cytotoxic drug is targeted to the tumor cells only. The drug is typically coupled to specific amino acids in the mAb chain,

with cysteine or lysine conjugation used most frequently. As most mAbs have a molecular weight of around 150 kDa and the drug molecule coupled with a linker moiety has a mass of a few hundred daltons, it is an analytical challenge to perform quality control, especially as one likes to know how many drug molecules are bound per mAb and whether they are equally distributed over all mAb molecules.

Cysteine-linked ADCs are composed of antibodies with drugs conjugated to cysteine residues involved in interchain disulfide links, *i.e.*, those between the two heavy chains (HCs) or those connecting the two light chains (LCs) to the HCs. Cysteine conjugation is performed via a prior partial reduction of the interchain disulfides, which implies that the HC-HC and LC-HC associations in the ADC monomer become a mixture of covalent and non-covalent associations. The potency of an ADC is determined in large part by the average number of drugs attached to the mAb, *i.e.*, the drug-to-antibody ratio (DAR). With each reduced disulfide bridge, two drug molecules can be coupled to the mAb, and thus, a typical cysteine-linked ADC carries two, four, six, or eight drug molecules. Valliere-Douglas *et al*.[375] used native MS early on to determine the intact masses of the non-covalently associated antibody HCs and LCs that result from the attachment of drug conjugates to interchain cysteine residues, focusing on IgG1 mAbs conjugated with maleimidocaproyl-monomethyl auristatin F (mcMMAF) or valine-citrulline-monomethyl auristatin E (vcMMAE). By using native MS on a high-resolution Bruker MaXis II Q-ToF mass spectrometer, they could indeed retain the intact bivalent structure of the ADC, which ordinarily would decompose as a consequence of the denaturing chromatographic conditions typically used for LC-MS analysis. In a follow-up study, the same group benchmarked SEC hyphenated to native MS versus analytical HIC to determine drug loads on a variety of ADCs[376]. The accurate results obtained were sufficient to support the use of native SEC-MS as a quantitative DAR method for all three tested ADCs. Similarly, Debaene *et al*.[377] explored the potential of native MS and IM-MS compared with HIC for quality control of the interchain cysteinyl-linked ADC brentuximab vedotin and obtained results in line with earlier data from Rosati *et al*.[117].

Dyachenko *et al*.[119] combined native MS with tandem MS for the characterization of brentuximab vedotin. Tandem MS allowed them to selectively fragment isolated precursors carrying a specific number of drug molecules. This enabled localization of cysteines to which the drug molecules were bound, revealing that drug conjugation took place nonhomogeneously to cysteine residues on both the LCs and HCs of the ADC when coupling was not saturated.

Next to cysteine-conjugated ADCs, lysine-conjugated ADCs have also been explored. Antibodies contain many more lysine residues than cysteines in their primary sequence, and this conjugation strategy therefore results in an even more heterogeneous ADC with potentially a much higher DAR. Conjugation to the free amines of the Lys residues depends on the solvent accessibility and reactivity of each individual Lys residue in the mAb. The overall lysine conjugation properties in human IgGs were studied by Gautier *et al*.[378], integrating high-resolution native MS and bottom-up proteomics. High-resolution Orbitrap native MS enabled monitoring of the sequential incorporation of up to 70 molecules, each attached to a different Lys, into human IgGs. Complementary bottom-up

proteomics facilitated the identification of the most reactive "hotspot" conjugation sites. Such data are important in controlling the drug load and specificity in lysine-linked ADCs.

For the next generation of ADCs, strategies that produce less heterogeneous drug loads are being explored. Illustratively, Botzanowski *et al*.[379] used native MS to investigate a site-specific DAR4 ADC generated through aldehyde-specific bioconjugation, whereby reactive formylglycine (fGly) amino acids are produced by formylglycine-generating enzyme (FGE) via highly selective oxidation of a cysteine residue found within a specific pentapeptide consensus sequence. The resulting mAb, containing four fGly residues, is then further modified using aldehyde-specific chemistries. The ADCs generated using these methods possess increased therapeutic indices and activities. By using online SEC hyphenated to native MS on a Q-ToF instrument, they observed a single species corresponding to a mass of 152,773 ± 1 Da, in agreement with the expected mass of the mAb with four conjugated drug molecules. Still, also a minor species (less than 10% of the total signal) was observed, corresponding to the binding of three drug molecules. These data clearly displayed the relatively high structural homogeneity of this mAb. Benchmarking their data, they analyzed in parallel the cysteine-conjugated mAb brentuximab vedotin and the lysine-conjugated mAb trastuzumab emtansine, and the native MS data clearly revealed the heterogeneity levels of these different classes of ADCs[379].

8.4. Other Biopharmaceuticals

Moving beyond antibody-based biotherapeutic molecules, high-resolution native MS has also been applied to study microheterogeneity in other biopharmaceuticals, also providing a means to distinguish between biosimilars and the original drug. Below we briefly highlight recent work on three commonly used non-antibody biopharmaceuticals: erythropoietin (EPO), etanercept, and human chorionic gonadotropin.

8.4.1. Erythropoietin

EPO is a human endogenous glycoprotein cytokine that is secreted mainly by the kidney in response to cellular hypoxia to stimulate red blood cell production (erythropoiesis). Recombinant EPO (rhEPO) is widely used in the treatment of anemia, for example in chronic kidney disease or myelodysplasia or as a result of cancer chemotherapy. Also, it is known that glycosylation of EPO affects its function in a variety of ways. For instance, sialylation and increased branching of EPO *N*-glycans increase its serum half-life[380], while EPO lacking sialylation exhibits a neuroprotective role in vivo[381]. Thus, a well-defined view of the detailed glycosylation profile of EPO is essential, also in the context of the fact that for rhEPO many biosimilars are reaching the market and clinic[382].

In one of the first studies using the Orbitrap EMR for detailed analysis of non-antibody-based biopharmaceuticals, Yang *et al*.[383] analyzed EPO. Compared with the relatively simple native MS spectra of therapeutic IgGs that typically display only a dozen of different glycoforms, high-resolution native mass spectra of EPO exhibited hundreds of different glycoproteoforms[383,384] arising from the heterogeneity on its one *O*- and three *N*-

glycosylation sites, making the compositional analysis of these spectra significantly more challenging. Treatment of EPO with a sialidase, which cleaves off all sialic acids present on both the *N*- and *O*-glycans on rhEPO, reduced the complexity of the spectra tremendously[383]. Combining the native MS analysis with a glycopeptide-centric MS approach[332] allowed them to assign most of the glycoforms. Additionally, they used the semiquantitative glycopeptide data to reconstruct the native mass spectra and looked for the overlap between the experimental and reconstructed native MS spectra[383]. Moreover, they analyzed the glycoproteoform profiles of three recombinant EPO products obtained from different manufacturers. These rhEPO therapeutics all had an identical protein backbone sequence, although they were found to be ornamented with differential glycosylation patterns, likely as a result of different conditions during their production. By directly comparing the glycoproteoform profiles obtained from the native mass spectra, Yang *et al*. introduced a biosimilarity score to describe the level of structural similarity between the three rhEPO therapeutics[383].

In a follow-up story, Čaval *et al*.[384] used a similar method for the structural characterization of 24 engineered variants of EPO with diverse *N*-glycosylation patterns. To this end, they employed a set of glycoengineered CHO cells to express EPO with different glycosylation features and degrees of heterogeneity[385]. Additionally, in that study the glycosylation profiles of the glycoengineered EPOs monitored by native MS revealed hundreds of co-occurring variants. Pleasingly, the chosen glycoengineering approach could be validated, revealing a stepwise decrease in heterogeneity from tetra-antennary polyLacNAc-elongated *N*-glycans all the way to homogeneous, diantennary, non-elongated, and non-sialylated *N*-glycans. Similar to Yang *et al*.[383], they employed a classification scheme based on the acquired native MS spectral fingerprints to create a product similarity matrix (**Figure 22**)[384]. Using hierarchical clustering, they were able to differentiate and classify glycoengineered EPO clones, extracting the overall structural differences. This classification scheme, termed biosimilarity scoring, could benefit clonal selection in cell line development and be used for batch-to-batch quality control of the glycoproteoform profiles and assessment of structural aspects linked to biosimilarity.

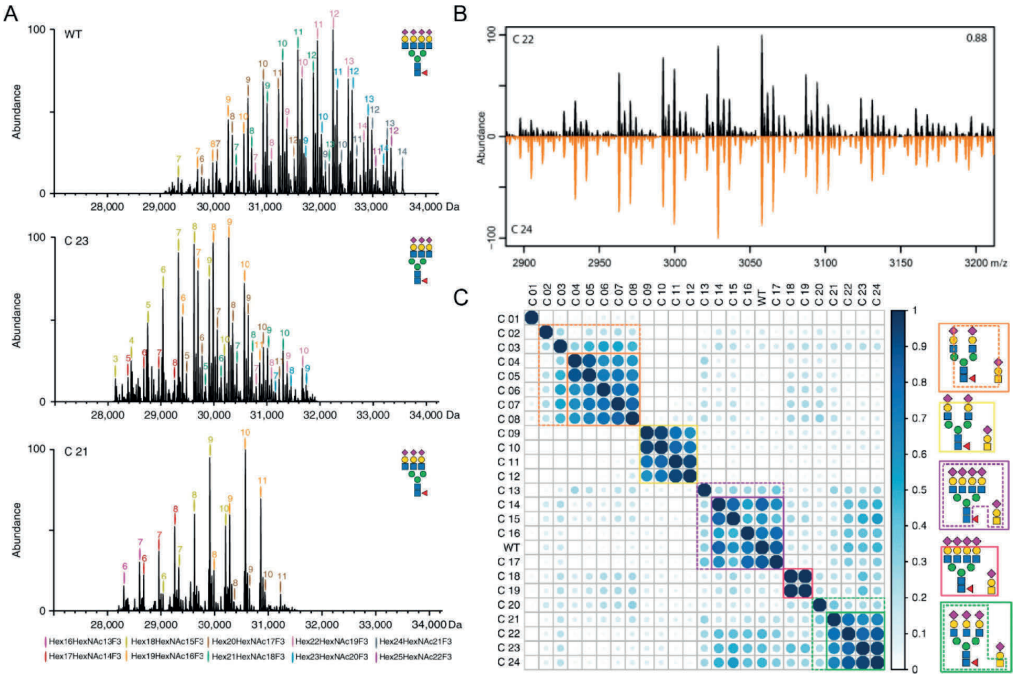


Figure 22 | Characterization of wild-type EPO and 23 glycoengineered EPO variants by high-resolution native MS. (A) Illustrative deconvoluted native mass spectra of WT EPO (top) and the knockout-based glycoengineered clones C23 (middle) and C21 (bottom). **(B)** Comparison of native mass spectra of EPO from two biological replicate clones C22 (black) and C24 (orange) by correlation scoring. **(C)** Clustering of the glycoengineered EPO clones based on the correlation between their native MS spectra. From ref [384]. CC BY 4.0.

8.4.2. Etanercept (Enbrel)

Combining high-resolution mass spectrometry with a sophisticated strategy of enzymatic digestions, Wohlschlager *et al.*[386] tackled the molecular complexity of the fusion protein etanercept (Enbrel). Etanercept consists of two copies of the tumor necrosis factor-α receptor (TNFR) and the Fc portion of human IgG1. This biopharmaceutical, which is used to treat autoimmune diseases such as rheumatoid arthritis, contains four *N*-glycosylation sites and 26 *O*-glycosylation sites in its dimeric TNFR domain and two *N*-glycosylation sites in the IgG Fc portion. Wohlschlager *et al.* used the protease IdeS to cleave etanercept at the hinge region of its IgG Fc portion, enabling separate analyses of the *N*-glycosylation on the TNFR and Fc domains independently. Next, the heterogeneity in glycosylation was trimmed by using in parallel and consecutively the three enzymes sialidase, PNGase F, and *O*-glycosidase. Moreover, glycopeptide-centric approaches were used concomitantly, and all data was integrated with advanced software[358] to achieve a comprehensive characterization of this extremely complex glycoprotein biopharmaceutical (**Figure 23**).

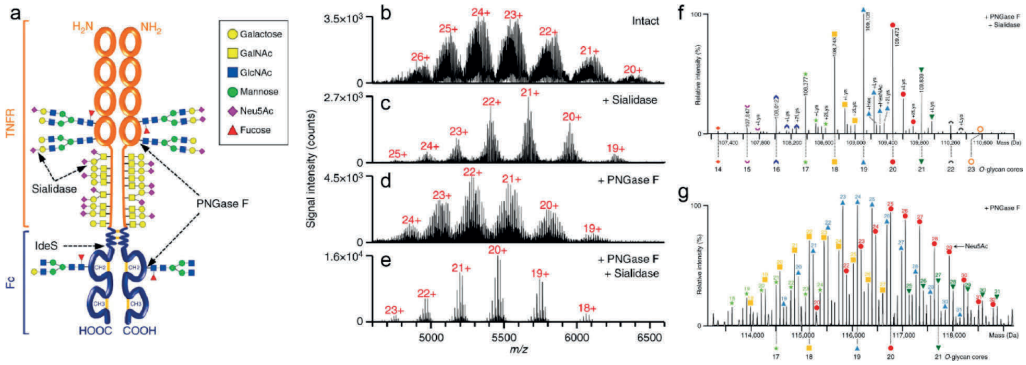


Figure 23 | Molecular structure and high-resolution native mass spectra of etanercept. (a) Structure of dimeric etanercept consisting of a TNFR and an Fc domain. Some of the cleavage sites of IdeS, PNGase F, and sialidase are indicated. **(b-e)** Native mass spectra of (b) intact etanercept and etanercept treated with (c) sialidase, (d) PNGase F, or (e) a combination of PNGase F and sialidase. **(f, g)** Annotation of the etanercept O-glycoforms after treatment with (f) PNGase F and sialidase or (g) PNGase F alone. Adapted from ref [386]. CC BY 4.0.

8.4.3. Human Chorionic Gonadotropin (Ovitrelle)

In an elegant combination of glycoproteoform profiling and non-covalent mass spectrometry, Lebede *et al.*[387] used high-resolution native MS to analyze recombinant human chorionic gonadotropin (hCG, Ovitrelle). The protein hormone hCG is involved in early embryo-maternal communication and maintenance of pregnancy. The two non-covalently interacting subunits, hCGα and hCGβ, both contain several *N*- and *O*-glycosylation sites, making the characterization of the proteoform profiles of the heterodimer even more challenging. Still, by integrating the native MS data with data from released glycans and glycopeptide analysis as well as intact mass analysis of the monomeric subunits and the heterodimer, they were able to provide an in-depth analysis of the heavily glycosylated non-covalent hCG heterodimer, suggesting that hCG is built up from more than 1000 distinct glycoforms. This is another example revealing that a single biopharmaceutical can in fact represent a highly diverse molecular landscape.

8.5. Proteoform Diversity of Selected Plasma Glycoproteins

At the molecular level, human plasma is a highly complicated biofluid. Its proteome exhibits an extraordinary dynamic range, making it hard to monitor lower-abundance proteins. About a dozen proteins represent 90% of the serum proteome in total protein concentration, and adding a dozen more increases this number to about 99%. In high-throughput proteomics experiments, these highly abundant proteins are often immune-depleted from plasma or serum, benefiting proteome coverage[388], although this depletion may also induce biases due to off-target codepletion. Staggering improvements in MS-based proteomics technologies over the past decade have renewed the interest in plasma proteomics, whereby now without depletion simultaneous quantitative monitoring of over 1000 plasma proteins can be achieved[389,390]. Interestingly, many established or

claimed plasma protein biomarkers, such as α-1-antitrypsin, C-reactive protein, ceruloplasmin, and haptoglobin, belong to the category of highly abundant proteins in plasma[391], and do not directly require these advances in proteomic depth.

In recent years, a few groups have started to use high-resolution native MS to especially target the most abundant plasma proteins with the aim of determining the variation in proteoforms present in individual donors. Such analyses provide a new dimension to the diversity of the plasma proteome[51]. Many of these abundant plasma proteins are glycoproteins (*e.g.*, α-1-antitrypsin, α-1-antichymotrypsin, ceruloplasmin, hemopexin, haptoglobin, α2-HS-glycoprotein, and various complement factors). They are predominantly expressed in the liver and subsequently secreted into the bloodstream. Changes in the glycosylation machinery of the liver cells that produce these proteins will directly translate to changes in the glycoproteome profiles of the resulting serum proteins, and therefore, plasma glycan and glycosylation analysis has become a rich source of potential biomarkers for diseases such as cancer[392,393] and chronic obstructive pulmonary disease (COPD)[394]. Although the glycosylation pattern of a plasma glycoprotein is largely determined by the glycosylation machinery of the (liver) cells in which they are produced, other factors may influence these proteoform profiles as well. For instance, liver cells may have biases in the glycoproteoforms that they secrete, and the lifetimes of plasma glycoproteins may be different for different glycoforms. Therefore, it is essential to monitor the plasma glycoproteoform profiles directly from plasma.

In some of the earlier work on human plasma, the proteins were purified from pooled plasma from multiple donors or obtained from commercial sources. However, already in those studies it was revealed that various plasma glycoproteins can exhibit a plethora of proteoforms, all having different molecular weights and potentially different functions. For instance, for complement protein C9, a striking number of at least 50 different proteoforms were exposed that had variations in *N*-, *O*-, and *C*-glycosylation[332,383]. Although this may suggest the presence of a very diverse glycoproteome profile, Franc *et al*.[395] concluded almost the opposite and stated that C9 exposed strong preferences for the amount and nature of glycans attached. Similar data were reported for the related proteins properdin[383] and C8, although the latter is somewhat more complicated because it is built up from three different chains (*i.e.*, C8α, C8β, and C8γ)[396].

8.5.1. Acute-Phase Proteins

Wu *et al*.[397,398] used high-resolution native MS to analyze purified α1-acid glycoprotein (AGP). AGP is a highly abundant acute-phase plasma protein that functions as a carrier for many hormones, lipids, and exogenous drugs and is proposed to deliver drug molecules to cells. The sources of the proteins were commercial and therefore likely originated from more than one donor. They were able to nicely disentangle the microheterogeneity of AGP, even though it harbors five *N*-glycosites. As the native mass spectra of sialylated AGP turned out to be very complex because of the varying degree of sialylation across all sites, they analyzed primarily asialo-AGP (neuraminidase-treated AGP) to facilitate glycan assignments. The resulting well-resolved and glycan-annotated native MS spectra of AGP are shown in **Figure 24**. Subsequently, they used native MS to measure binding of warfarin

to AGP, which revealed that different glycoproteoforms of AGP exhibit distinct affinities for the drug. Most notably, an increase both in fucosylation and *N*-glycan branching/elongation reduced the binding to warfarin. Similarly, *N*-glycan branching and elongation also decreases binding of asialo-AGP to warfarin.

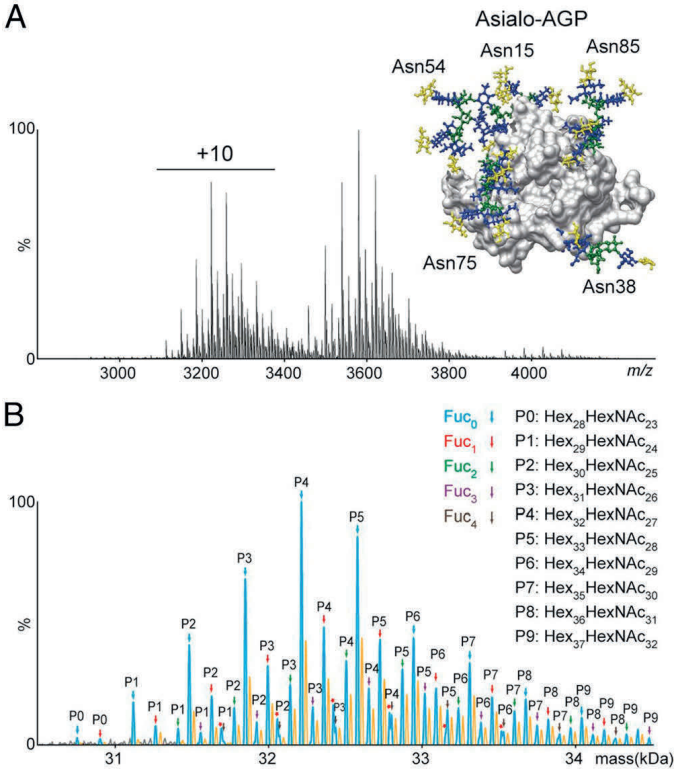


Figure 24 | Glycan-annotated high-resolution native mass spectra of α1-acid glycoprotein (AGP). (A) High-resolution native mass spectrum and **(B)** zero-charged and glycan annotated deconvoluted mass spectrum of asialo-AGP. The inset in (A) shows the AGP structure with five highly branched *N*-glycans at Asn15, Asn38, Asn54, Asn75, and Asn85. From ref [397]. CC BY-NC-ND 4.0.

A breakthrough towards personalized plasma proteome profiling came when Lin *et al.*[399] demonstrated that roughly the top 30 most abundant plasma proteins can be efficiently purified from individual donors using only 50 µL of plasma for analysis by high-resolution native MS. They subjected serum samples to various forms of prefractionation, notably SEC and various forms of IEX. The purification of individual proteins from individual donors opened up the interesting question whether each donor exhibits a unique proteoform profile for each serum glycoprotein[399]. This question was first tackled by analyzing α2-HS-glycoprotein (AHSG, fetuin) extracted from serum from 20 different donors, of which 10 had experienced a septic episode. Lin *et al.*[399] observed that the proteoform profile of fetuin for each of these donors turned out to be unique and thus highly personalized. This diversity could be explained to a great extent by the presence of two dominant alleles of fetuin in the population (AHSG1 and AHSG2), which could be

either homozygous or heterozygous. The mutations present in AHSG also affect an important *O*-glycosylation site, with mutation of a Ser into a Thr leading to a highly abundant Thr *O*-glycosylation, whereas the Ser was found to be mostly unmodified. Although the proteomic differences between donors of different genotypes were already interesting, also donors of the same genotype exhibited striking differences in phosphorylation and fucosylation, whereby the latter seemed to be enhanced in the older patients and especially those that had experienced a septic episode.

Using partly the same cohort of donors, Čaval *et al.*[400] used high-resolution native MS to profile the proteoforms of intact alpha-1-antichymotrypsin (AACT). This protein was purified from individual plasma samples that were longitudinally collected from 10 patients at four time points over the duration of a septic episode. As determined by standard proteomics, AACT, a positive acute-phase protein, followed a similar abundance profile as C-reactive protein, being elevated during sepsis but returning back to baseline when the patients were dismissed. More interestingly, the proteoform profiles of AACT from the 10 patients gradually increased in mass over the monitored period, which could be ascribed to increased levels of fucosylation, antennary branching, and LacNAc moieties in response to the septic episode (**Figure 25**). This glycoproteoform remodeling thus extended over a much longer period than the septic episode. Although several trends in AACT abundance and extent of glycosylation were alike in all 10 monitored patients, it was apparent that each donor exhibited a unique glycoproteoform profile.

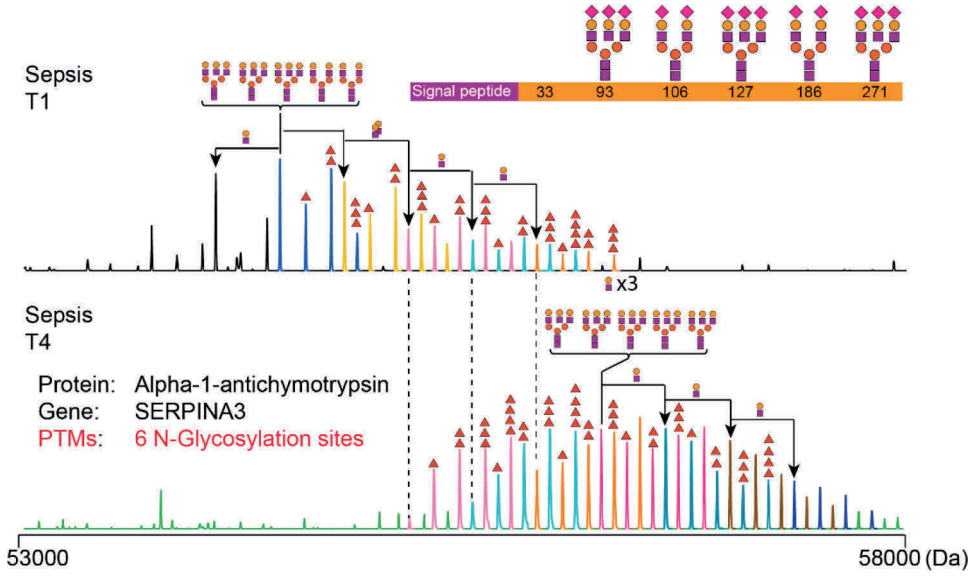


Figure 25 | Native-MS-derived proteome profiles of serum SerpinA3 (AACT) from a donor before and after a septic episode revealing an extensive increase in glycosylation. From ref [400].

8.5.2. Haptoglobin

Both the Robinson and Heck groups studied the abundant plasma protein haptoglobin (Hp) by high-resolution native MS. Haptoglobin has been implicated as a biomarker for several pathological conditions[401,402], but in most studies, merely the abundance level of haptoglobin is considered as the biomarker. In humans, there are two frequently occurring allelic forms of Hp, resulting in three major genotypes: homozygous Hp 1-1 and Hp 2-2 and heterozygous Hp 2-1. This genetic polymorphism has an intriguing effect on the quaternary structure of Hp. The simplest genotype, Hp 1-1, yields dimers consisting of two α1β units connected by disulfide bridges. Hp 2-1 forms mixtures of linear $(\alpha1)_2(\alpha2)_{n-2}(\beta)_n$ oligomers (n > 1) while Hp 2-2 occurs in cyclic $(\alpha2)_n(\beta)_n$ oligomers (n > 2). The main function of Hp in plasma is that it scavenges toxic hemoglobin (Hb) leaked into the bloodstream from erythrocytes. It has been shown that different Hp genotypes bind Hb with different affinities, with Hp 2-2 being the weakest binder.

Hp also harbors four *N*-glycosylation sites on the β-chain, located at Asn184, Asn207, Asn211, and Asn241, all occupied by complex type *N*-glycans with a varying number of antennas, which may be fucosylated and sialylated[403,404]. Wu *et al*. combined high-resolution native MS and glycoproteomics to investigate Hp 1-1 glycosylation microheterogeneity and examine its impact on the interactions with Hb and lectins [397,398]. Already the ~ 80 kDa Hp 1-1 dimers exhibited quite complex glycoproteoform profiles. Using affinity chromatography, they established that *N*-glycan branching attenuates Hp-Hb binding and, on the contrary, that fucosylation stabilizes Hp-Hb binding. Overall, these studies indicated that *N*-glycosylation fine-tunes Hp-Hb interactions. Building further upon these findings, Tamara *et al*.[405] analyzed Hp from all three genotypes: homozygous Hp 1-1 and Hp 2-2 and heterozygous Hp 2-1. They first combined native MS and mass photometry to study the oligomer distributions in all three genotypes qualitatively and quantitatively. Subsequently, they dissected the glycoproteoform profiles of individual oligomer states using a combination of SEC and IEX with high-resolution native MS and obtained well-resolved glycoproteoform profiles for – among others – the 138, 188, and 237 kDa Hp 2-1 trimer, tetramer, and pentamer, respectively (**Figure 26**). These high-resolution native mass spectra were matched with simulations based on quantitative intact-mass LC-MS data for distinct Hp subunits generated in parallel, ultimately facilitating glycoproteoform annotations. Strikingly, these high-resolution native MS spectra revealed that each oligomer displayed distinct glycosylation patterns. Subsequently, also the hemoglobin binding propensities of these distinct oligomers were probed by affinity chromatography, revealing that Hb binding is tightly and finely regulated by both Hp oligomerization and Hp glycosylation. Overall, these studies clearly revealed that a wealth of genotype-specific proteoforms fine-tunes hemoglobin scavenging for haptoglobin in plasma, which should also be considered when designing cell-free haptoglobin-based therapeutics[406].

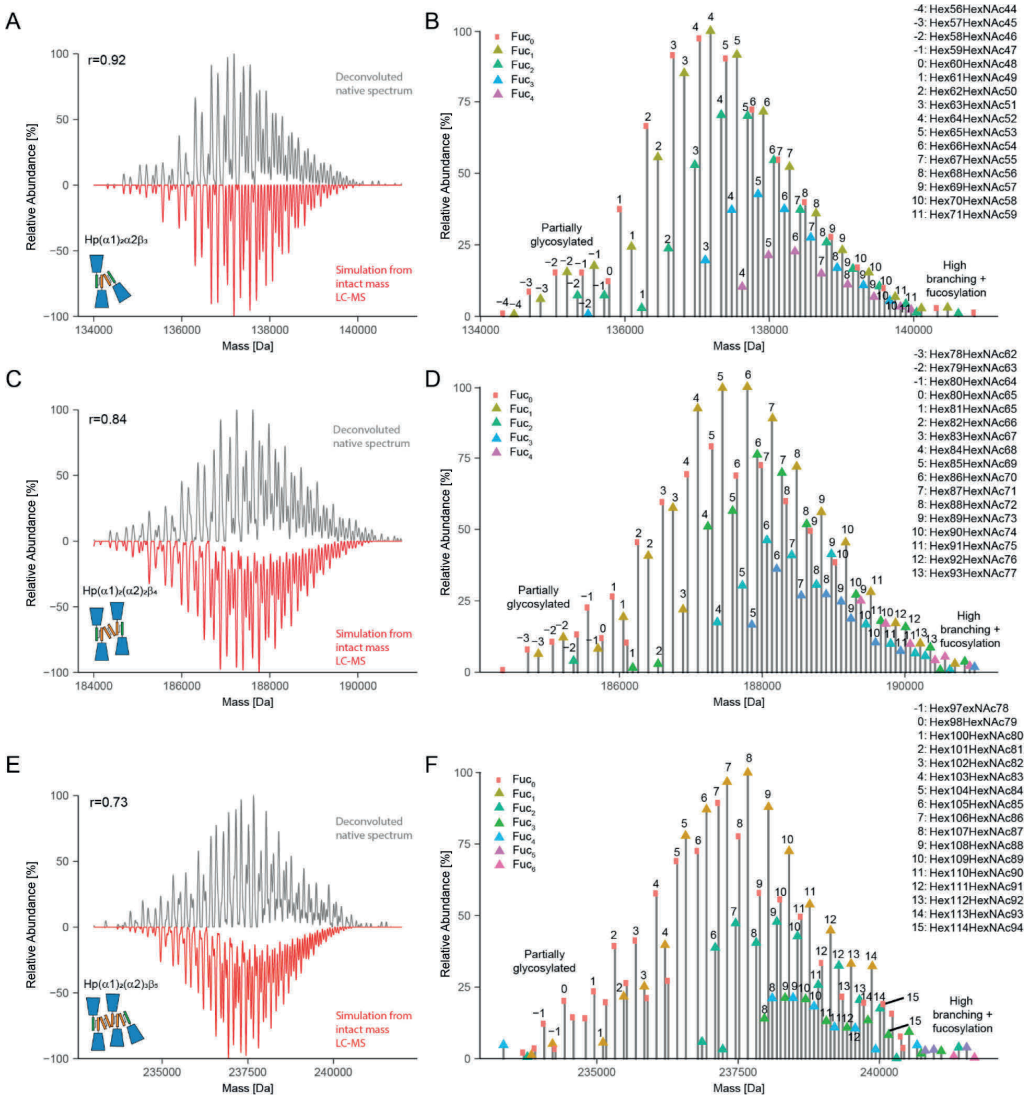


Figure 26 | Comparison of deconvoluted high-resolution native mass spectra with simulated mass spectra reconstructed from intact-mass hydrophilic interaction chromatography (HILIC) LC-MS data. (A, C, E) Mirror plots for Hp 2-1 multimers: (A) $(\alpha1)_2\alpha2\beta_3$ trimer, (C) $(\alpha1)_2(\alpha2)_2\beta_4$ tetramer, and (E) $(\alpha1)_2(\alpha2)_3\beta_5$ pentamer. The zero-charged deconvoluted native MS spectrum is depicted at the top and the in silico-reconstructed spectrum at the bottom. The Pearson correlation coefficient (*r*) for each pair of native and in silico-reconstructed spectra is given in the top left corner of each panel, and a schematic drawing of each Hp 2-1 multimer is shown in the bottom left corner of each panel. **(B, D, F)** Annotation of peaks detected in the native deconvoluted MS spectra for the (B) trimer, (D) tetramer, and (F) pentamer of Hp 2-1. From ref [405]. CC BY-NC-ND 4.0.

These initial studies focusing on just a single serum glycoprotein have provided a new dimension of quantitative plasma proteome profiling. For almost every plasma glycoprotein, the glycosylation profile seems to be unique for each individual donor in

terms of proteoforms and their abundances. This striking observation opens the way to a next level of personalized proteome profiling in which the unique proteoform profiles are used to stratify patient cohorts and to monitor their adaptation to physiological changes. More research is needed to find out which plasma protein(s) will be the best biomarker(s) for each particular disease, but it is apparent that glycoproteoform profiles need to be considered in future plasma proteomics.

8.6. High-Resolution Native MS of Intact Phosphoproteins

In the previous section, we described how high-resolution native MS can be very powerful for the analysis of protein glycosylation. However, this technology is equally applicable to the analysis of intact proteins decorated with other PTMs. Protein phosphorylation is an important modification that is involved in signal transduction, may regulate protein-protein interactions, and (de)activates kinases and phosphatases, among others. The ability of native MS to investigate protein-protein interactions and protein phosphorylation concomitantly was exploited by Kleppe *et al.*[407] to study the phosphorylation dependence and stoichiometry of the complex formed by tyrosine hydroxylase and 14-3-3γ. Tyrosine hydroxylase (TH) is a high-affinity binding partner of 14-3-3γ protein, but this interaction is tightly regulated by phosphorylation. Kleppe *et al*. performed native MS analyses of human TH (non-phosphorylated or phosphorylated on Ser19 (TH-pS19) or Ser40 (TH-pS40) alone and together with 14-3-3γ. Whereas tetrameric TH-pS19 (224 kDa) bound 14-3-3γ (58.3 kDa) with high affinity (K_d = 3.2 nM), generating complexes containing either one (282.4 kDa) or two (340.8 kDa) dimers of 14-3-3γ, no complex formation between non-phosphorylated TH or TH-pS40 and 14-3-3γ was observed, highlighting the crucial role of pS19.

Similarly, van de Waterbeemd *et al.*[408] used a combination of high-resolution native MS and bottom-up proteomics to monitor the effects of protein phosphorylation on complex formation. More in-depth, in this study the phosphorylation and cyclic nucleotide binding of dimeric 150 kDa cGMP-dependent protein kinase (PKG) were simultaneously monitored by high-resolution native MS, which showed that binding of cAMP or cGMP causes different PKG phosphorylation kinetics. In a second example, it was demonstrated that the binding and phosphorylation of the mitotic regulator Bora by the cycle kinase Aurora A proceed independently. Interestingly, the three investigated proteins – Aurora A, Bora, and PKG – existed in different phosphorylation states. The relative abundances of these phosphoisoforms could be accurately monitored by native MS, whereas complementary peptide-centric MS experiments were done to localize the phosphorylated residues[408]. In an extension to those experiments, high-resolution native MS was also used by Lössl *et al.*[409] to decipher the phosphoproteoforms of Bora resulting from phosphorylation by either Aurora A or Polo-like kinase 1 (Plk1), showing that the two kinases target different Bora residues and generate distinct phosphorylation cascades. The tripartite Aurora A/Bora/Plk1 interplay modulates the cell's recovery from DNA-damage-induced cell cycle arrest, and it is characterized by numerous mutual phosphorylation events with various implications for protein structure, interactions, and function. Lössl *et al*. simultaneously probed these often temporarily occurring effects by

integrating native MS, cross-linking-MS, IM-MS, top-down sequencing, and bottom-up proteomics[409]. In a somewhat related study, Abdul Azeez *et al*.[410] utilized native MS and IM-MS to determine the phosphorylation-dependent formation of a complex between Aurora C and the inner centromere protein (INCENP). They introduced mutations on different Ser residues of the TSS motif, which weakened the interaction substantially. IM-MS data revealed that the phosphorylated Aurora C-INCENP complex exhibited higher cross sections, hinting at a more flexible structure containing partly disordered regions.

In another elegant study, Ben-Nissan *et al*.[70] used high-resolution native (tandem) MS to analyze the yeast homotetrameric ~ 155 kDa FBP1 complex. FBP1 is the rate-limiting enzyme in gluconeogenesis. The data revealed that each of the subunits in the FBP1 tetramer is differently phosphorylated when expressed under different growth conditions, whereby the subsequent incorporation of each of the four phosphate moieties (80 Da each) could be nicely resolved. In this manner, they were able to determine the stoichiometry, the kinetics, and by top-down proteomics the exact position of the phosphorylations. Similarly, Potel *et al*.[411] used high-resolution native (tandem) MS to monitor the stoichiometry and phosphorylation in heterohexamers of the tumor metastasis suppressor Nm23. Human Nm23 is present in various isoforms, of which Nm23-H1 and Nm23-H2 are by far the most dominant. Extracting these hexameric assemblies from different compartments of different cells, they demonstrated cell- and compartment-specific stoichiometries of this abundant complex. Subsequently, they combined native and top-down MS to investigate the histidine autophosphorylation activity of the purified Nm23 assemblies.

9. Non-covalent Interactions of Proteins with Small Ligands Probed by High-Resolution Native MS

Interactions between proteins and small ligands are fundamental to many cellular activities and processes. A wide variety of small molecules, including metal ions, heme groups, metabolites, hormones, drugs, (oligo)nucleotides, and peptides, can non-covalently bind to proteins to regulate their structure and function. Understanding and quantifying such protein-ligand interactions is therefore of key importance for biological research and the discovery and design of effective therapeutics. The affinities and kinetics of protein-ligand interactions are typically assessed by various biochemical and biophysical techniques, the most common of which are surface plasmon resonance (SPR)[412,413], circular dichroism (CD)[414], and isothermal titration calorimetry (ITC)[415]. Using native MS to study protein-ligand interactions seems a very attractive prospect due to the technique's inherent simplicity and sensitivity, as well as the richness of the information that can be obtained. Native MS does not require chemical labeling or immobilization, and sample consumption is relatively low. With direct mass measurements, one can also resolve multiple coexisting species, determine the stoichiometry of ligand binding, and investigate cooperativity.

The first studies into protein-ligand interactions by native MS were initiated in the early 1990s with the works of Ganem, Li and Henion[10,416] and Katta and Chait[11] shortly after the

introduction of ESI. Attempts at quantitative assessment of affinities quickly followed suit, with Loo and coworkers providing some of the first examples of such studies[417]. Because of the unique ability of native MS to monitor the abundance of coexisting species simultaneously, several different strategies emerged. Affinities could be determined from a titration experiment by following the intensity ratio between the protein and the protein-ligand complex[418-420]. Alternatively, a secondary ligand of known affinity was used as a reference point in competitive binding experiments[418-420], with Jørgensen *et al*.[418] being among the first to report absolute values for dissociation constants as determined by native MS. The field received another impulse with the advent of automation[421], as Zhang *et al*.[422] introduced the NanoMate, a combination of an autosampler with a chip-based nano ESI source that could be used to drastically increase the throughput, making the technology ready for application in the pharmaceutical industry. By the early 2000s, several reviews discussed this new role of native MS as a tool for quantitative studies of protein-ligand interactions[423-425]. It was expected that measurements of dissociation constants would have been routine by now, especially with the availability of automation. However, a lack of consistency in early literature led to a "hit or miss" perception, limiting the use to a few specialist laboratories. For native MS to become a more widespread technique for studies of protein-ligand binding, several questions needed to be answered, both fundamental and practical in nature, as discussed in the following sections.

9.1. Optimizing Native MS for the Quantitative Determination of Binding

With appropriate testing and optimization, quantitative information about ligand binding can be obtained from native MS experiments, even though it is not an inherently quantitative technique. This relates mostly to the ESI process but also to artifacts that may be introduced by ion transfer and detection in the gas phase. One of the main questions raised early on was whether the ESI process would distort the solution-phase equilibrium. Measurement times should be kept short, as slow acidification of the sample may occur through electrochemical water oxidation in positive ion mode[426,427]. Acidification also occurs within the ESI droplets themselves, but this is thought to happen on a time scale that is too short to impact the equilibrium substantially[428,429]. Aqueous ammonium acetate solutions only partially alleviate this issue, as their buffering capacity only becomes meaningful when the pH drops towards 4.75, the pK_a of acetic acid[430]. According to the charge residue model, evaporation also leads to a short-lived increase in analyte concentrations. Although too short to affect the equilibrium substantially, this can still introduce nonspecific binding, especially when high concentrations of protein and ligands are used. However, it has been shown that such weakly bound species can often easily be removed further downstream in the mass spectrometer[424].

Several studies have focused on how ionization efficiency, ion transfer, and detection may differentially affect different analytes. Although an equal response in ionization would be a reasonable assumption when the species have similar sizes and biochemical properties (*e.g.*, non-covalent protein complexes that differ by a single bound small ligand), this is often not the case for metal-binding, DNA/RNA-binding, and larger protein complexes.

The ionization efficiency depends not only on the charge but also on access to the surface of the ESI droplet, which has been shown to vary with the size, conformation, and hydrophobicity of the analyte[424,431]. Furthermore, complexes stabilized by ionic and/or polar interactions are much more stable in the gas phase than those stabilized by hydrophobic interactions[432-434]. A response factor accounting for ionization efficiency and instrumental bias was introduced by Gabelica *et al*.[435], with Tjernberg *et al*.[436] shortly thereafter also addressing gas-phase dissociation. Alternatively, the approximation of response factors can be avoided altogether by including a reference complex with a known K_d value, for instance as described by Kempen and Brodbelt[437].

Native MS also differs from solution-based techniques in that it is known to be quite incompatible with solvents containing nonvolatile salts. High metal salt concentrations, which are often desirable to preserve protein-ligand equilibria, result in protein ions carrying multiple adducts, which not only dilutes the ion signal but also complicates the detection of low-abundance protein-ligand complexes. Partly as a result of this, researchers have often struggled to detect full ligand binding by native MS for interactions dependent on the presence of metal ions[422]. However, in recent studies[438-440] this problem has been largely circumvented using nanoscale ion emitters. In contrast with micrometer spraying capillaries, the narrower-diameter capillaries produce even smaller droplets that carry fewer nonvolatile components. In a comprehensive work, Nguyen *et al*.[438] demonstrated that a narrower tip diameter (~ 250 nm) can help to improve the determination of ligand-protein binding affinities. For example, they obtained quite accurate K_d values for binding of an inhibitor to the metalloenzyme carbonic anhydrase (**Figure 27**).

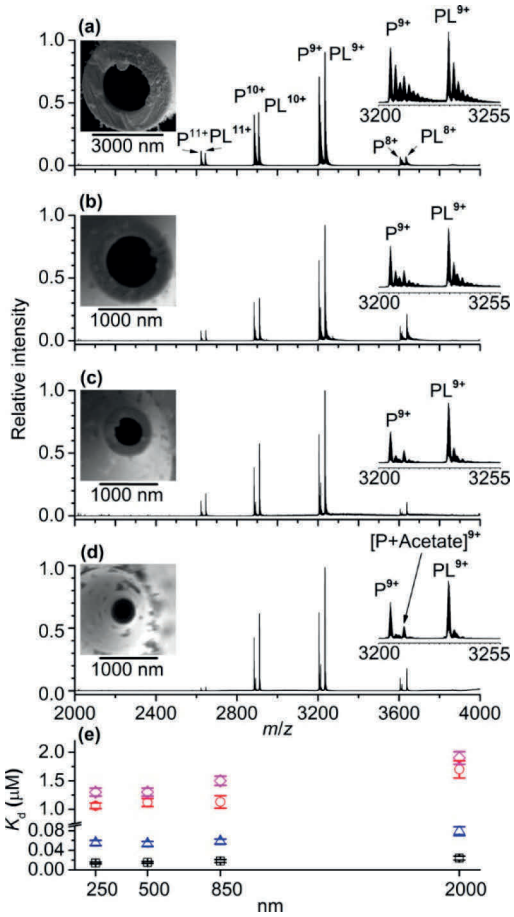


Figure 27 | Narrow-bore nanoelectrospray ionization emitters with inner tip diameters less than 1000 nm can be used to obtain more accurate K_d values for carbonic anhydrase (P) inhibitors (L). (a-d) nESI mass spectra of aqueous solutions containing 5 μM human carbonic anhydrase I, 2 μM ethoxzolamide, and 70 mM ammonium acetate obtained using emitter tips with inner diameters of (a) ~ 2000, (b) ~ 850, (c) ~ 500, and (d) ~ 250 nm. **(e)** K_d values measured using nESI as a function of the emitter tip diameter for the binding of ethoxzolamide (squares), brinzolamide (circles), furosemide (triangles), and dichlorphenamide (diamonds) to human carbonic anhydrase I. Reproduced from ref [438]. Copyright 2019 American Chemical Society.

Although native MS can thus be used for the quantitative assessment of protein-ligand binding constants, there are limitations in the range of K_d values that can be measured, also depending on the type of experiment. With the most standard titration assays, the ion intensity ratio between the free protein and the protein-ligand complex is measured. By deducing the corresponding free ligand concentration or by measuring it directly using a calibration curve, one can make a Scatchard plot, from which the K_d value can be extracted. The dynamic range of this approach is limited, however, as the target protein concentration should ideally be below the expected K_d value while the ligand concentration is titrated through this range. In practice, this means that only affinities above 100 nM can be measured reliably by this native MS approach. On the other hand, competitive binding experiments can be applied in which a secondary ligand of known

affinity is used to compete with a high-affinity ligand of interest. This effectively shifts the binding curve toward higher concentrations that can be measured by native MS. In this manner, nanomolar-range affinities can be measured, even though typical native MS concentrations in the micromolar range are employed in these experiments. By focusing on the free ligand in the lower m/z range of the instrument, Wortmann *et al*.[441] showed that the lower limit of K_d measurements could be extended even below the picomolar range.

With careful optimization of experimental parameters to avoid biases and artifacts, native MS can thus provide reliable quantitative data on protein-ligand binding. Systematic comparisons between ESI-MS and other methods by the groups of Klassen, Zenobi, and others have shown that native MS has a unique position within the toolbox for quantitative assessment of protein-ligand interactions[442-448]. Notably, Jecklin *et al*.[447] investigated a set of eight well-characterized inhibitors of human carbonic anhydrase I (hCAI) with K_d values spanning over 4 orders of magnitude (low nM to high μM). While good agreement was seen for four inhibitors, limitations of the different techniques became apparent when the others were analyzed. While the assessment by native MS was found to be hampered by gas-phase dissociation of hydrophobic ligands and nonspecific adducts at higher concentrations, native MS was able to overcome issues related to fast kinetics (hampering SPR) and poor solubility (hampering ITC). This study by Jecklin *et al*. clearly reveals that every method may provide false positives and false negatives, and thus, cross-validation of approaches should always be encouraged.

9.2. Pharmaceutical Applications of Protein-Ligand Screening by Native MS

Mass spectrometry is nowadays used in the early stages of drug discovery to screen compounds or fragment libraries against a protein target in fairly high-throughput approaches[449-454]. The role of native MS in fragment-based lead discovery (FBLD) was pioneered by Swayze *et al*.[455], who identified structure-activity relationships (SARs) to generate a lead compound binding the 1061 region of bacterial 23S rRNA. Hits from a first fragment screen revealed two interesting SAR trends, and when these hits were combed, ternary complexes observed in native MS readily showed that these two trends involved different binding sites. Competitive binding experiments on ligand derivatives provided additional information on how the two binding sites may be combined in rationally synthesized fused compounds. The resulting lead compounds displayed up to a 20-fold increase in affinity as well as increased functional activity. An automated approach was presented a few years later by Maple *et al*.[456], who used the NanoMate platform to screen a library of 157 compounds against an apoptotic protein target. Although extensive optimization of the instrumental parameters was required, both the throughput and results proved to be comparable to those of NMR- or ITC-based library screening. A similar strategy employed by Woods *et al*.[457] using a 720-member fragment library also showed a very good correlation between native MS and traditional screening setups based on SPR and X-ray crystallography. Moreover, these and other studies[458] showed that sample consumption and the required fragment concentrations were lower in the MS-based

approach, widening accessibility to poorly soluble compounds that would otherwise have been discarded. More recently, an alternative automation approach was presented by Ren *et al.*[459], using an autosampler and SEC to remove unbound compounds from a target protein incubated with pools of small molecules followed by native MS to improve the sensitivity (**Figure 28**).

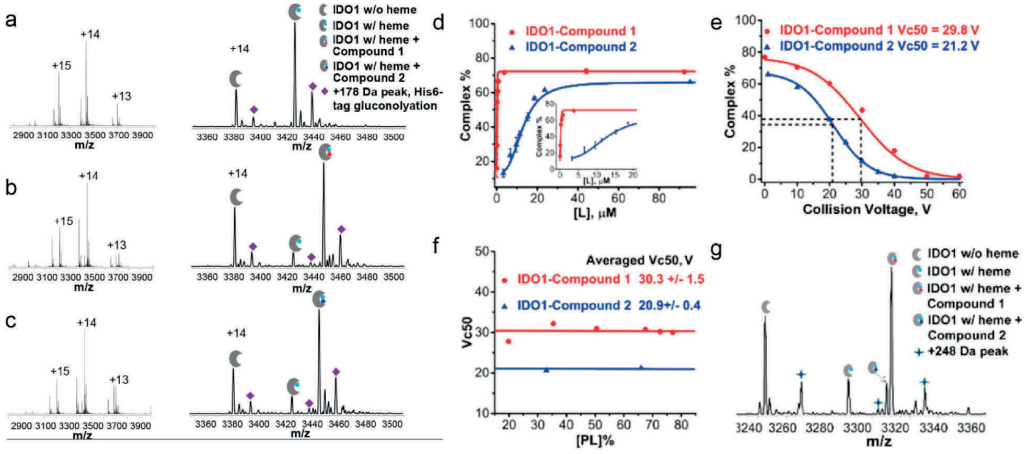


Figure 28 | Automated screening of protein-ligand interactions using a compound library and native SEC-MS. (a-c) Native mass spectra of the enzyme indoleamine-pyrrole 2,3-dioxygenase (IDO1) alone, IDO1 with compound 1, and IDO1 with compound 2, respectively. The left panels show the charge-state distributions, and the right panels focus on the enlargements of the +14 charge state for identification. For all of the spectra shown, the concentration of IDO1 was 5 µM (2 µM for IDO1 w/heme), whereas the concentration of compound 1 was 3.5 µM in (b) and that of compound 2 was 10 µM in (c). **(d)** Saturation curves from titration experiments. The concentration of the IDO1 sample was 13 µM (9 µM for IDO1 w/heme). The concentrations of compounds 1 and 2 ranged from 0.5 to 100 µM. **(e)** In-source dissociation profiles for the IDO1-compound 1 and IDO1-compound 2 complexes. Dashed lines indicate the corresponding Vc_{50} values. The concentration of the IDO1 sample was 12 µM, and the concentrations of compounds 1 and 2 were 89 µM. **(f)** Plot of Vc_{50} against initial complex percentage at different compound 1/compound 2 to IDO1 ratios. **(g)** Competition experiment for evaluation of relative binding affinities. IDO1 (total 13 µM, IDO1 w/heme 9 µM) was incubated with compounds 1 and 2 (each at 10 µM). Each precursor experiences a different lab-frame energy based on its charge state. All of the data points in (a) and (b) were replicated three times, with error bars shown on the graph. Adapted from ref [459]. Copyright 2018 American Chemical Society.

The advantages and complementarity of ligand screening by native MS also become apparent in the later stages of drug development, including the analysis of high-affinity compounds in the hit-to-lead (H2L) and lead optimization (LO) processes[449,452,454]. Chip-based native MS can be used in competitive binding experiments for affinity ranking of optimized compounds. Bovet *et al.*[460] demonstrated this principle using different ligands binding the human estrogen receptor, although poor reproducibility prevented them from providing reliable K_d values for hydrophobic interactions, which are typically weakened in the gas phase. Another competitive binding study by Jecklin *et al.*[461] compared a set of clinical kinase inhibitors and introduced a chip-based method focusing solely on the unbound ligands. Better sensitivity and resolution in the low-m/z range circumvented the need for harsher ion transmission conditions – and therefore dissociation. Both applied approaches, focused on either intact ligand-protein complexes or unbound ligands, led to

results that were in good agreement with known IC_{50} values. Evidently, for the latter approach, focused solely on analyzing the free ligand, high resolution is easily achieved even on conventional mass spectrometers.

9.3. Binding of Small Molecules to Heterogeneous Proteins Analyzed by High-Resolution Native MS

Outside of a few laboratories equipped with high-resolution FT-ICR mass spectrometers, studying protein-ligand binding was often limited by the mass resolution of the typically employed ToF instruments. With the advent of the Orbitrap mass spectrometers as a more accessible high-resolution MS platform, protein-ligand binding studies became possible for smaller and smaller ligands in interaction with more heterogeneous larger proteins. In a follow-up study to their earlier work, Maple *et al*.[462] were among the first to employ the Orbitrap EMR instrument for high-throughput ligand screening. Baseline separation of glycoforms enabled the approach to be extended toward glycoproteins not tractable by traditional ToF-based approaches. Furthermore, the sensitivity advantage of the Orbitrap instrument enabled Maple *et al*. to measure proteins at lower concentrations, assess stronger interactions, and reduce sample consumption.

Following up on their pioneering work in making native MS amenable for the analysis of membrane proteins[90], the Robinson group started to use Orbitraps with extended mass range to probe small-molecule and lipid binding to this important class of proteins in more detail. In a breakthrough study, Gault *et al*.[21] measured a wide range of membrane proteins on an Orbitrap instrument, using the HCD cell to strip off residually bound detergent molecules. Because of the high resolving power of the instrument, unambiguous distinctions could be made between lipids from the same class with different chain lengths and degrees of saturation when tightly bound to the trimeric OmpF assembly (**Figure 29**). Moreover, analysis of the interaction between OmpF and a small peptide revealed three successive binding events for which the K_d could be determined individually. A follow-up study by Mehmood *et al*.[463] used the same approach on the human intramembrane zinc metalloprotease ZMPSTE24 to reveal and quantitate the off-target binding and activity of several HIV protease inhibitors. In 2018, Gupta *et al*.[235] released a protocol for identifying lipids that tightly bind to membrane protein complexes and explored how delipidation affects the disruption of oligomeric interactions.

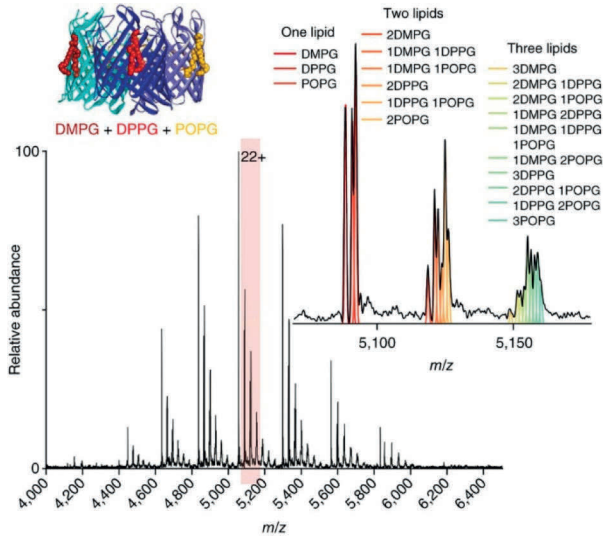


Figure 29 | Native MS spectrum (main panel) of trimeric OmpF bound to an equimolar ratio of DMPG, DPPG, and POPG lipids (cartoon inset). The 22+ charge state is shown in an expanded view (right), with peaks showing up to three bound lipids. Theoretical distributions corresponding to different combinations of lipids are shown by colored lines and correlate with the spectrum. Reproduced with permission from ref [21]. Copyright 2016 Nature Publishing Group.

9.4. Resolving Multiple Binding Events to Oligomeric Protein Assemblies and DNA/RNA Molecules

Whereas concomitant binding to multisubunit complexes cannot be accurately assessed by standard approaches for probing protein-ligand binding (*e.g.*, ITC, SPR), native MS can distinguish by mass and thus reveal the entire distribution of ligand-bound states. Gavriilidou *et al.*[464] demonstrated this unique advantage of native MS in their study of the dimer-tetramer equilibrium of M2 Pyruvate Kinase (PKM2), a regulatory enzyme that is often inactive in the glycolytic pathway in tumor cells. An allosteric activator, fructose-1,6-bisphosphate (FBP), was found to shift the dimer-tetramer equilibrium toward the active tetramer, with the 4:4 stoichiometry of FBP binding to the tetramer only. Another study by Root *et al.*[465] investigated the isoprenoid biosynthesis enzyme IspF from *Arabidopsis thaliana*, a homotrimeric assembly that binds multiple ligands, including a metal cofactor and a synthetic inhibitor. Whereas standard biophysical techniques failed to reveal the mode of action of recently discovered inhibitors, native MS enabled researchers to propose a mechanism that involves competition with the substrate and extraction of the Zn^{2+} from the active site. Using native MS, El-Baba *et al.*[466] recently also investigated the ability of various small molecules to modulate the activity of dimeric SARS-CoV-2 main protease, including a compound proposed to disrupt the oligomeric interface.

Apart from analyzing protein-ligand interactions, native MS provides a great means for in-depth investigation of the interactions between DNA/RNA and small molecules. In a comprehensive screening study, Gülbakan *et al.*[467] captured all major aspects of aptamer-

ligand interactions for three DNA-aptamers, showing the stoichiometry, selectivity, and cooperativity of various interactions. They supplemented MS data with SPR, ITC, and CD measurements and highlighted the unique strengths of native MS for aptamer-ligand analysis, whereby the different techniques complemented each other. Recently, Nguyen *et al*.[468] applied native MS and its ability to distinguish between multiple binding modes to study how netrospin, a potent antibiotic and anticancer agent precursor, interacts with hairpin and duplex DNA molecules. Using nanoscale emitters, they were able to spray DNA-ligand samples with high salt concentrations, enabling them to simultaneously determine binding affinities for five ligand-DNA and DNA-DNA complexes.

9.5. Investigating the Kinetics and Thermodynamics of Ligand Binding

An additional attractive feature of native MS for the assessment of protein-ligand interactions is that the abundance of different species can be followed over the course of the measurement, enabling the investigation of kinetics and thermodynamics[469,470]. Using a temperature-controlled ESI source, Cong *et al*.[471] investigated the thermodynamics of lipid binding to AmtB, an integral membrane protein of *Escherichia coli*. Their approach allowed them to determine the thermodynamics of individual binding events for lipids with variable chain length, resolving unique thermodynamic signatures. In another report, Moghadamchargari *et al*.[472] demonstrated that high-resolution native MS could be used to study the intrinsic inactivation rates of the oncoprotein K-RAS and mutants thereof. This was done by monitoring the hydrolysis of non-covalently bound GTP, as identified by a mass shift corresponding to the loss of a phosphate group (**Figure 30**). The inactivation rates were in good agreement with complementary methods measuring the concentration of released organic phosphate. However, native MS unexpectedly revealed that some oncogenic mutants have a higher intrinsic hydrolysis rate for 2′-deoxy-GTP than for GTP, which are separated in mass by only 16 Da. Another approach explored by Marchand *et al*.[473] determined the entropic and enthalpic contributions to the binding equilibrium of G-quadruplex nucleic acid structures and their ligands using a temperature-controlled nESI source.

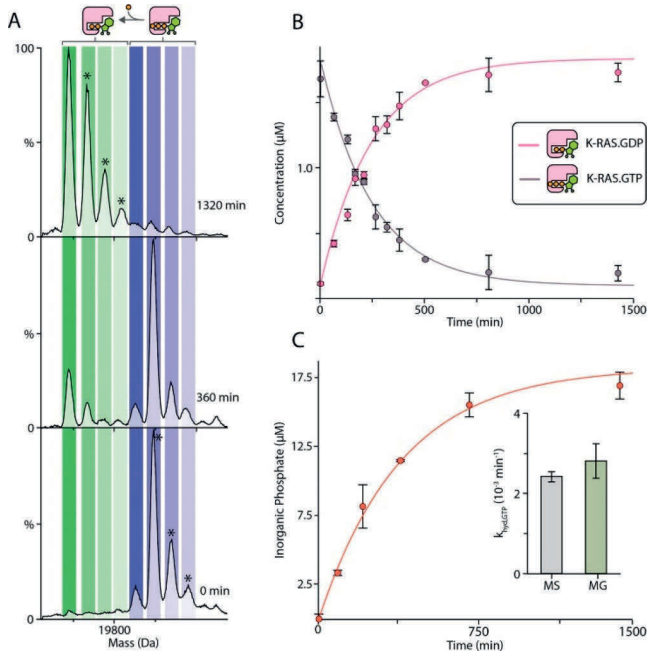


Figure 30 | Determination of the guanosine triphosphate (GTP) hydrolysis rate ($k_{hyd,GTP}$) for K-RAS. (A) Representative native mass spectra recorded at different time points for K-RAS (2 μM) loaded with GTP incubated at 25 °C. The abundance of K-RAS bound to guanosine diphosphate (GDP) increases as GTP is hydrolyzed. Asterisks denote species with bound sodium or magnesium adducts. **(B)** The concentration of K-RAS bound to GDP and GTP was determined from deconvolution of the native MS data (dots) and fitted to a first-order rate constant model (solid lines). **(C)** Plot of inorganic phosphate concentration determined for K-RAS·GTP using a malachite green assay. The inset shows the similar rate constants determined by native MS and the MG assay. Reproduced from ref [472]. Copyright 2019 American Chemical Society.

9.6. Determination of the Ligand-Binding Site Using Native Top-Down ECD and UVPD Fragmentation

High-resolution native MS provides additional advantages for the study of protein-ligand interactions in that it can probe the ligand-binding interfaces. Using fast and high-energy ion activation techniques, the protein backbone can be fragmented, and this may occur without breaking the non-covalent interactions. Such dissociation results in a mass shift corresponding to the ligand's mass, ideally for both the N- and C-terminal fragments. The idea of probing interaction sites in this manner was pioneered by Xie *et al.*[474], who used ECD on α-synuclein in complex with polycationic spermine. This allowed them to localize the binding site to residues 106-138, consistent with data obtained by NMR spectroscopy. The approach is applicable to both electrostatic interactions that are stable in the gas phase[475,476] and hydrophobic interactions that are more prone to dissociation[477]. Several of the first applications of UVPD to native protein-ligand complexes also revealed that ligands could remain bound to the fragments[223,225,478,479]. For example, O'Brien *et al.*[225] demonstrated that the C-terminal fragment ions of peptidyl-prolyl cis-trans isomerase 1 (Pin1) that retained a peptide ligand were formed by UVPD within the functionally

relevant WW domain. The only N-terminal fragments that retained the peptide originated from the PPIase domain. This was in agreement with the reported crystal structure. These studies demonstrate some of the seemingly ever-expanding opportunities that are open for researchers by bridging native MS with a multitude of activation methods, both established and just emerging in the field of native top-down MS.

10. Conclusions and Outlook

Since its introduction by John Fenn, electrospray ionization has impacted biomolecular mass spectrometry way beyond the early high expectations. Whole new research fields such as proteomics and metabolomics now use mass spectrometry as their core technology and combine advanced separation technologies with electrospray ionization. Deservedly, Fenn was awarded the Nobel Prize for this invention in 2002, which he accepted with a Nobel Lecture entitled; "Electrospray Wings for Molecular Elephants"[480]. At that time, early on in this century, it was slowly becoming apparent that these flying molecular elephants could be partly kept "alive" while flying through the vacuum of the mass spectrometer on their way to the detector for mass analysis. To achieve this, the solvent conditions in the electrospray process needed to be adjusted, avoiding organic solvents and acidification, typically used in LC-MS[15]. Primarily by using volatile buffer-mimicking solutions, such as aqueous ammonium acetate, primary, secondary, tertiary, and even quaternary structures of proteins and protein complexes could be largely retained, leading to the advent of what is now known as native mass spectrometry[14]. By now, about 20 years later, native MS has matured into a versatile analytical method, with widespread applications in molecular and structural biology and beyond in all areas of the life sciences[3,20]. In this review, we have broadly categorized the applications into (1) analysis of protein assemblies, (2) analysis of proteoforms, and (3) analysis of protein-ligand interactions, but evidently, many of the applications reported here and in other literature cross over these three categories. Furthermore, many other applications are emerging and maturing, such as the analysis of DNA and RNA molecules, ribonucleoprotein complexes, and membrane-embedded protein complexes.

In this review, we have presented published work describing what is already possible by high-resolution native MS, but at the end of this review it is tempting to ask what is not yet possible? Even more, what do we dream of?

A dream likely shared by many would be to perform "sample-preparation-free" native MS analysis. We have described here the laborious efforts often needed to purify protein complexes of interest, aiming to desalt these analytes as much as possible, for instance by several rounds of dialysis, while simultaneously avoiding the copurification of detrimental detergent molecules and polymers. However, with the increased sensitivity and capabilities to desolvate ions within the latest mass analyzers[65,70,75], several pioneering attempts have been made to analyze proteins and protein complexes directly from cellular broths[481] or cellular membranes[324], reasonable success has already been achieved. Such attempts will certainly expand and make native MS hopefully even more feasible also for nonexperts.

Another futuristic dream is to be able to count and measure the mass of every molecule and molecular assembly within a cell. This is quite a challenge, as it has been estimated that there are ~ 10^9 protein molecules[482] in a given human cell, each with its own specific function, interactions, and potential decoration with a unique set of PTMs. Charting individual molecules requires single-molecule approaches. Several single-molecule analytical approaches have been developed over the past decades[483-485], but mass spectrometry has lagged behind and largely remained an ensemble-based method. However, several single-molecule and single-particle mass spectrometry approaches are now emerging[108,127,149] and may go hand-in-hand with other technologies that can be used to measure masses of single particles, such as mass photometry[257] and nanomechanical-systems-based mass analysis[486]. Single-particle approaches enhance not only the sensitivity but also the specificity, as each molecule can be analyzed by itself, revealing its own unique features. This will surely help to chart unknown territories, such as those presented by the wide variety of proteoforms co-occurring within a cell. Potentially, one day we will be able to assert that each protein within the cell is indeed unique. Ideally, these two dreams can be combined, enabling each molecule within the cell to be measured with minimal distortions and sample preparation.

High-resolution mass spectrometry, as powerful as it is, delivers even more biological information when used in concert with other technologies. It has taken some adjustments, but as reviewed also here, native MS has now become an integral part of many structural biology studies, contributes to the exciting fields of protein design and engineering, and has already contributed to COVID-19-related science[339,340]. Native mass spectrometry has benefited from the resolution revolution in electron microscopy[487,488], but also, vice versa, sample preparation for electron microscopy can be assisted by prior native MS analysis[489-492]. While high-resolution structural methods provide unprecedented detail toward structure and function, native mass spectrometry is very well suited to reveal the molecular heterogeneity inherent in each and every biological sample. Within the discipline of mass spectrometry, native MS is more and more conjugated with hydrogen-deuterium exchange MS, cross-linking MS, chemical labeling, and interaction proteomics[3]. From a niche technology, it has now become a key technology embedded in the structural biology toolbox.

The field of high-resolution native mass spectrometry still has many challenges ahead. Through a concerted effort with parallel future developments in mass analyzers, sample preparation, separation, and data analysis, we can dream on and expect fast evolution of native MS in the near future.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

Since its conception about 30 years ago, the research field of native MS has grown steadily, which by now has led to a rather substantial and lively research community. We acknowledge all involved, making this a scientifically competitive but very friendly and cooperative community. We acknowledge all researchers whose work has made it possible for us to write this comprehensive review. We apologize for missing contributions made by others; the field has grown so much that it is rather difficult to be all-inclusive. We thank all members of the Heck lab for their support in writing this review and for executing some of the work described in it, and we also thank all our collaborators over the years for supplying interesting research questions and precious samples. A.J.R.H. acknowledges support from The Netherlands Organization for Scientific Research (NWO) through the Spinoza Award (SPI.2017.028). M.A.d.B. and A.J.R.H. were additionally supported by NWO NACTAR Project 16442.

References

1. Altelaar, A.F.M., Munoz, J. & Heck, A.J.R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* **14**, 35-48 (2013).
2. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).
3. Lössl, P., van de Waterbeemd, M. & Heck, A.J.R. The diverse and expanding role of mass spectrometry in structural and molecular biology. *The EMBO Journal* **35**, 2634-2657 (2016).
4. Masson, G.R. et al. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595-602 (2019).
5. Zheng, J., Strutzenberg, T., Pascal, B.D. & Griffin, P.R. Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry. *Curr. Opin. Struct. Biol.* **58**, 305-313 (2019).
6. Li, K.S., Shi, L. & Gross, M.L. Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc. Chem. Res.* **51**, 736-744 (2018).
7. O'Reilly, F.J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25**, 1000-1008 (2018).
8. Liu, F. & Heck, A.J.R. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35**, 100-108 (2015).
9. Iacobucci, C., Gotze, M. & Sinz, A. Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr. Opin. Biotechnol.* **63**, 48-53 (2020).
10. Ganem, B., Li, Y.T. & Henion, J.D. Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray mass spectrometry. *Journal of the American Chemical Society* **113**, 7818-7819 (1991).
11. Katta, V. & Chait, B.T. Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* **113**, 8534-8535 (1991).
12. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
13. Leney, A.C. & Heck, A.J.R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28**, 5-13 (2017).
14. Heuvel, R.H.H.v.d. & Heck, A.J.R. Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8**, 519-526 (2004).
15. Loo, J.A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1-23 (1997).
16. Heck, A.J.R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927-933 (2008).

17. Sharon, M. & Robinson, C.V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76**, 167-193 (2007).
18. Sharon, M., Witt, S., Glasmacher, E., Baumeister, W. & Robinson, C.V. Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome. *J. Biol. Chem.* **282**, 18448-18457 (2007).
19. Bakhtiari, M. & Konermann, L. Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *J. Phys. Chem. B* **123**, 1784-1796 (2019).
20. Robinson, C.V. Mass spectrometry: From plasma proteins to mitochondrial membranes. *Proc Natl Acad Sci U S A* **116**, 2814-2820 (2019).
21. Gault, J. et al. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods* **13**, 333-336 (2016).
22. Kebarle, P. & Verkerk, U.H. Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **28**, 898-917 (2009).
23. Wilm, M. Principles of Electrospray Ionization. *Mol. Cell. Proteom.* **10** (2011).
24. Banerjee, S. & Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012**, 282574 (2012).
25. Konermann, L., Ahadi, E., Rodriguez, A.D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **85**, 2-9 (2013).
26. Garza, K.Y. et al. Desorption Electrospray Ionization Mass Spectrometry Imaging of Proteins Directly from Biological Tissue Sections. *Anal. Chem.* **90**, 7785-7789 (2018).
27. Rose, R.J., Damoc, E., Denisov, E., Makarov, A. & Heck, A.J.R. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* **9**, 1084-1086 (2012).
28. Li, H., Nguyen, H.H., Ogorzalek Loo, R.R., Campuzano, I.D.G. & Loo, J.A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **10**, 139-148 (2018).
29. Zhang, H., Cui, W., Wen, J., Blankenship, R.E. & Gross, M.L. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal. Chem.* **83**, 5598-5606 (2011).
30. Fort, K.L. et al. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. *Analyst* **143**, 100-105 (2018).
31. Rosati, S. et al. Exploring an Orbitrap Analyzer for the Characterization of Intact Antibodies by Native Mass Spectrometry. *Angew. Chem. Int. Ed.* **51**, 12992-12996 (2012).
32. Lössl, P., Snijder, J. & Heck, A.J.R. Boundaries of Mass Resolution in Native Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **25**, 906-917 (2014).
33. Rathore, D. et al. The role of mass spectrometry in the characterization of biologic protein products. *Expert Rev. Proteom.* **15**, 431-449 (2018).
34. Snijder, J., Rose, R.J., Veesler, D., Johnson, J.E. & Heck, A.J.R. Studying 18 MDa Virus Assemblies with Native Mass Spectrometry. *Angew. Chem. Int. Ed.* **52**, 4020-4023 (2013).
35. Wilkinson., C.b.A.D.M.a.A. IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). (Blackwell Scientific Publications, Oxford, 1997).
36. Murray, K.K. et al. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry* **85**, 1515-1609 (2013).
37. Xian, F., Hendrickson, C.L. & Marshall, A.G. High resolution mass spectrometry. *Anal. Chem.* **84**, 708-719 (2012).
38. Marshall, A.G., Hendrickson, C.L. & Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**, 1-35 (1998).
39. Pelander, A., Decker, P., Baessmann, C. & Ojanperä, I. Evaluation of a High Resolving Power Time-of-Flight Mass Spectrometer for Drug Analysis in Terms of Resolving Power and Acquisition Rate. *J. Am. Soc. Mass Spectrom.* **22**, 379-385 (2011).
40. Tsybin, Y.O., Nagornov, K.O. & Kozhinov, A.N. in Fundamentals and Applications of Fourier Transform Mass Spectrometry 113-132 (Elsevier, 2019).
41. Makarov, A. et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78**, 2113-2120 (2006).
42. Guilhaus, M., Selby, D. & Mlynski, V. Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom. Rev.* **19**, 65-107 (2000).
43. Bowman, A.P. et al. Ultra-High Mass Resolving Power, Mass Accuracy, and Dynamic Range MALDI Mass Spectrometry Imaging by 21-T FT-ICR MS. *Anal. Chem.* **92**, 3133-3142 (2020).

44. Zuth, C., Vogel, A.L., Ockenfeld, S., Huesmann, R. & Hoffmann, T. Ultrahigh-Resolution Mass Spectrometry in Real Time: Atmospheric Pressure Chemical Ionization Orbitrap Mass Spectrometry of Atmospheric Organic Aerosol. *Anal. Chem.* **90**, 8816-8823 (2018).
45. Olsen, J.V. et al. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010-2021 (2005).
46. Marshall, A.G. & Rodgers, R.P. Petroleomics: Chemistry of the underworld. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18090-18095 (2008).
47. Smith, R.D. et al. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2**, 513-523 (2002).
48. Lebedev, A.T. Environmental mass spectrometry. *Annu. Rev. Anal. Chem.* **6**, 163-189 (2013).
49. Orr, A., Stotesbury, T., Wilson, P. & Stock, N.L. The use of high-resolution mass spectrometry (HRMS) for the analysis of DNA and other macromolecules: A how-to guide for forensic chemistry. *Forensic Chemistry* **14**, 100169 (2019).
50. Selliez, L. et al. High-resolution mass spectrometry for future space missions: Comparative analysis of complex organic matter with LAb-CosmOrbitrap and laser desorption/ionization Fourier transform ion cyclotron resonance. *Rapid Commun. Mass Spectrom.* **34** (2020).
51. Yang, Y., Franc, V. & Heck, A.J.R. Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends Biotechnol.* **35**, 598-609 (2017).
52. van de Waterbeemd, M. et al. Dissecting ribosomal particles throughout the kingdoms of life using advanced hybrid mass spectrometry methods. *Nat. Commun.* **9** (2018).
53. Kafader, J.O. et al. in Journal of the American Society for Mass Spectrometry, Vol. 31 574-581 (NLM (Medline), 2020).
54. Brenton, A.G. & Godfrey, A.R. Accurate mass measurement: Terminology and treatment of data. *J. Am. Soc. Mass Spectrom.* **21**, 1821-1835 (2010).
55. Gross, M.L. Accurate masses for structure confirmation. *J. Am. Soc. Mass Spectrom.* **5**, 57 (1994).
56. Dawson, P. Quadrupole mass spectrometry and its applications. (books.google.com, 2013).
57. Savory, J.J. et al. Parts-Per-Billion Fourier Transform Ion Cyclotron Resonance Mass Measurement Accuracy with a "Walking" Calibration Equation. *Anal. Chem.* **83**, 1732-1736 (2011).
58. Senko, M.W., Beu, S.C. & McLafferty, F.W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **6**, 229-233 (1995).
59. Hernández, H. & Robinson, C.V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2**, 715-726 (2007).
60. Gorshkov, M.V., Good, D.M., Lyutvinskiy, Y., Yang, H. & Zubarev, R.A. Calibration Function for the Orbitrap FTMS Accounting for the Space Charge Effect. *J. Am. Soc. Mass Spectrom.* (2010).
61. Makarov, A., Grinfeld, D. & Ayzikov, K. in Fundamentals and Applications of Fourier Transform Mass Spectrometry 37-61 (Elsevier, 2019).
62. Chernushevich, I.V. & Thomson, B.A. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* **76**, 1754-1760 (2004).
63. Snijder, J. & Heck, A.J.R. Analytical Approaches for Size and Mass Analysis of Large Protein Assemblies. *Annu. Rev. Anal. Chem.* **7**, 43-64 (2014).
64. Schachner, L.F. et al. Standard Proteoforms and Their Complexes for Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **30**, 1190-1198 (2019).
65. van de Waterbeemd, M. et al. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat. Methods* **14**, 283-286 (2017).
66. Shinholt, D.L., Anthony, S.N., Alexander, A.W., Draper, B.E. & Jarrold, M.F. A frequency and amplitude scanned quadrupole mass filter for the analysis of high m/z ions. *Rev. Sci. Instrum.* **85**, 113109 (2014).
67. Lippens, J.L. et al. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *Journal of the American Society for Mass Spectrometry* **29**, 183-193 (2018).
68. Laganowsky, A., Reading, E., Hopper, J.T.S. & Robinson, C.V. Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639-651 (2013).
69. Mamyrin, B.A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *Int. J. Mass spectrom.* **206**, 251-266 (2001).
70. Ben-Nissan, G. et al. Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex. *Anal. Chem.* **89**, 4708-4715 (2017).

71. Campuzano, I.D.G. et al. High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *J. Am. Soc. Mass Spectrom.* **31**, 1155-1162 (2020).
72. Shaw, J.B. et al. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *J. Am. Soc. Mass Spectrom.* **27**, 1929-1936 (2016).
73. Mallis, C.S. et al. Development of native MS capabilities on an extended mass range Q-TOF MS. *Int. J. Mass spectrom.* **458**, 116451 (2020).
74. Wyttenbach, T., Kemper, P.R., Baykut, G., Park, M.A. & Bowers, M.T. A new instrument with high mass and high ion mobility resolution. *Int. J. Mass spectrom.* **434**, 108-115 (2018).
75. Gault, J. et al. Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nat. Methods* **17**, 505-508 (2020).
76. Inoue, R. et al. New insight into the dynamical system of αb-crystallin oligomers. *Sci. Rep.* **6** (2016).
77. Cameron, A.E. & Eggers, D.F. An Ion ``Velocitron''. *Rev. Sci. Instrum.* **19**, 605-607 (1948).
78. Mamyrin, B., Karataev, V., Shmikk, D. & Zagulin, V. The mass-reflectron. A new nonmagnetic time-of-flight high resolution mass-spectrometer. *Zhurnal Ehksperimental'noj i Teoreticheskoj Fiziki* **64**, 82-89 (1973).
79. Mirgorodskaya, O.A., Shevchenko, A.A., Chernushevich, I.V., Dodonov, A.F. & Miroshnikov, A.I. Electrospray-ionization time-of-flight mass spectrometry in protein chemistry. *Anal. Chem.* **66**, 99-107 (1994).
80. Dawson, J.H.J. & Guilhaus, M. Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **3**, 155-159 (1989).
81. Krutchinsky, A.N., Zhang, W. & Chait, B.T. Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. *J. Am. Soc. Mass Spectrom.* **11**, 493-504 (2000).
82. Morris, H.R. et al. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **10**, 889-896 (1996).
83. Sobott, F., Hernández, H., Mccammon, M.G., Tito, M.A. & Robinson, C.V. A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies. *Anal. Chem.* **74**, 1402-1407 (2002).
84. Tahallah, N., Pinkse, M., Maier, C.S. & Heck, A.J.R. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun. Mass Spectrom.* **15**, 596-601 (2001).
85. Lorenzen, K., Versluis, C., Van Duijn, E., Van Den Heuvel, R.H.H. & Heck, A.J.R. Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int. J. Mass spectrom.* **268**, 198-206 (2007).
86. Van Den Heuvel, R.H.H. et al. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473-7483 (2006).
87. Uetrecht, C. et al. High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9216-9220 (2008).
88. Uetrecht, C. et al. Subunit exchange rates in Hepatitis B virus capsids are geometry- and temperature-dependent. *Phys. Chem. Chem. Phys.* **12**, 13368 (2010).
89. Shoemaker, G.K. et al. Norwalk Virus Assembly and Stability Monitored by Mass Spectrometry. *Mol. Cell. Proteom.* **9**, 1742-1751 (2010).
90. Barrera, N.P. & Robinson, C.V. Advances in the Mass Spectrometry of Membrane Proteins: From Individual Proteins to Intact Complexes. *Annu. Rev. Biochem.* **80**, 247-271 (2011).
91. Barrera, N.P. et al. Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat. Methods* **6**, 585-587 (2009).
92. Fischer, P. & Schweikhard, L. Multiple-ion-ejection multi-reflection time-of-flight mass spectrometry for single-reference mass measurements with lapping ion species. *Rev. Sci. Instrum.* **91**, 023201 (2020).
93. Knauer, S. et al. A multi-reflection time-of-flight setup for the improvement and development of new methods and the study of atomic clusters. *Int. J. Mass spectrom.* **446** (2019).
94. Dziekonski, E.T., Johnson, J.T., Lee, K.W. & McLuckey, S.A. Fourier-Transform MS and Closed-Path Multireflection Time-of-Flight MS Using an Electrostatic Linear Ion Trap. *Anal. Chem.* **89**, 10965-10972 (2017).

95. Chernushevich, I.V., Merenbloom, S.I., Liu, S. & Bloomfield, N. A W-Geometry Ortho-TOF MS with High Resolution and Up to 100% Duty Cycle for MS/MS. *J. Am. Soc. Mass Spectrom.* **28**, 2143-2150 (2017).
96. Richardson, K. & Hoyes, J. A novel multipass oa-TOF mass spectrometer. *Int. J. Mass spectrom.* **377**, 309-315 (2015).
97. Comisarow, M.B. & Marshall, A.G. Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **25**, 282-283 (1974).
98. Hendrickson, C.L. et al. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **26**, 1626-1632 (2015).
99. Li, H., Wolff, J.J., Van Orden, S.L. & Loo, J.A. Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **86**, 317-320 (2014).
100. Comisarow, M.B. & Marshall, A.G. Theory of Fourier transform ion cyclotron resonance mass spectroscopy. I. Fundamental equations and low-pressure line shape. *J. Chem. Phys.* **64**, 110-119 (1976).
101. Bartholdi, E. & Ernst, R.R. Fourier spectroscopy and the causality principle. *J. Magn. Reson.* **11**, 9-19 (1973).
102. Marshall, A.G. & Hendrickson, C.L. High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.* **1**, 579-599 (2008).
103. Senko, M.W., Hendrickson, C.L., Emmett, M.R., Shi, S.D.H. & Marshall, A.G. External accumulation of ions for enhanced electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **8**, 970-976 (1997).
104. Marshall, A.G. & Guan, S. Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **10**, 1819-1823 (1996).
105. Marty, M.T. et al. Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes. *Anal. Chem.* **84**, 8957-8960 (2012).
106. Zhang, H., Cui, W., Wen, J., Blankenship, R.E. & Gross, M.L. Native Electrospray and Electron-Capture Dissociation in FTICR Mass Spectrometry Provide Top-Down Sequencing of a Protein Component in an Intact Protein Assembly. *J. Am. Soc. Mass Spectrom.* **21**, 1966-1968 (2010).
107. Campuzano, I.D.G. et al. Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **90**, 745-751 (2018).
108. Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **17**, 395-398 (2020).
109. Zubarev, R.A. & Makarov, A. Orbitrap mass spectrometry. *Anal. Chem.* **85**, 5288-5296 (2013).
110. Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156-1162 (2000).
111. Kingdon, K.H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* **21**, 408-418 (1923).
112. Hu, Q., Cooks, R.G. & Noll, R.J. Phase-enhanced selective ion ejection in an orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **18**, 980-983 (2007).
113. Hu, Q., Makarov, A.A., Cooks, R.G. & Noll, R.J. Resonant ac Dipolar Excitation for Ion Motion Control in the Orbitrap Mass Analyzer. *J. Phys. Chem. A* **110**, 2682-2689 (2006).
114. Lange, O., Damoc, E., Wieghaus, A. & Makarov, A. Enhanced Fourier transform for Orbitrap mass spectrometry. *Int. J. Mass spectrom.* **369**, 16-22 (2014).
115. Makarov, A., Denisov, E. & Lange, O. Performance evaluation of a high-field orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **20**, 1391-1396 (2009).
116. Snijder, J. et al. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *J. Am. Chem. Soc.* **136**, 7295-7299 (2014).
117. Rosati, S. et al. In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *MAbs* **5**, 917-924 (2013).
118. Belov, M.E. et al. From protein complexes to subunit backbone fragments: A multi-stage approach to native mass spectrometry. *Anal. Chem.* **85**, 11163-11173 (2013).
119. Dyachenko, A. et al. Tandem Native Mass-Spectrometry on Antibody–Drug Conjugates and Submillion Da Antibody–Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **87**, 6095-6102 (2015).

120. Campuzano, I.D.G. et al. Native MS Analysis of Bacteriorhodopsin and an Empty Nanodisc by Orthogonal Acceleration Time-of-Flight, Orbitrap and Ion Cyclotron Resonance. *Anal. Chem.* **88**, 12427-12436 (2016).
121. Snyder, D.T., Panczyk, E.M., Somogyi, A., Kaplan, D.A. & Wysocki, V. Simple and minimally invasive SID devices for native mass spectrometry. *Anal. Chem.* **92**, 11195-11203 (2020).
122. Mehaffey, M.R., Sanders, J.D., Holden, D.D., Nilsson, C.L. & Brodbelt, J.S. Multistage Ultraviolet Photodissociation Mass Spectrometry to Characterize Single Amino Acid Variants of Human Mitochondrial BCAT2. *Anal. Chem.* **90**, 9904-9911 (2018).
123. Greisch, J.-F. et al. Expanding the mass range for UVPD-based native top-down mass spectrometry. *Chem. Sci.* **10**, 7163-7171 (2019).
124. Poltash, M.L., McCabe, J.W., Shirzadeh, M., Laganowsky, A. & Russell, D.H. Native IM-Orbitrap MS: Resolving what was hidden. *Trends Analyt. Chem.* **124** (2020).
125. Poltash, M.L. et al. Fourier Transform-Ion Mobility-Orbitrap Mass Spectrometer: A Next-Generation Instrument for Native Mass Spectrometry. *Anal. Chem.* **90**, 10472-10478 (2018).
126. Kafader, J.O. et al. Measurement of Individual Ions Sharply Increases the Resolution of Orbitrap Mass Spectra of Proteins. *Anal. Chem.* **91**, 2776-2783 (2019).
127. Keifer, D.Z. & Jarrold, M.F. Single-molecule mass spectrometry. *Mass Spectrom. Rev.* **36**, 715-733 (2017).
128. Qi, Y. et al. Absorption-Mode: The Next Generation of Fourier Transform Mass Spectra. *Anal. Chem.* **84**, 2923-2929 (2012).
129. Grinfeld, D., Aizikov, K., Kreutzmann, A., Damoc, E. & Makarov, A. Phase-constrained spectrum deconvolution for fourier transform mass spectrometry. *Anal. Chem.* **89**, 1202-1211 (2017).
130. Kelstrup, C.D. et al. Limits for Resolving Isobaric Tandem Mass Tag Reporter Ions Using Phase-Constrained Spectrum Deconvolution. *J. Proteome Res.* **17**, 4008-4016 (2018).
131. Allison, T.M. & Landreh, M. in Comprehensive Analytical Chemistry, Vol. 83 161-195 (Elsevier B.V., 2019).
132. Göth, M. & Pagel, K. Ion mobility–mass spectrometry as a tool to investigate protein–ligand interactions. *Anal. Bioanal. Chem.* **409**, 4305-4310 (2017).
133. López, A., Tarragó, T., Vilaseca, M. & Giralt, E. Applications and future of ion mobility mass spectrometry in structural biology. *New J. Chem.* **37**, 1283-1289 (2013).
134. Österlund, N., Moons, R., Ilag, L.L., Sobott, F. & Graslund, A. Native ion mobility-mass spectrometry reveals the formation of β-barrel shaped amyloid-β hexamers in a membrane-mimicking environment. *J. Am. Chem. Soc.* **141**, 10440-10450 (2019).
135. May, J.C. et al. Conformational landscapes of ubiquitin, cytochrome c, and myoglobin: Uniform field ion mobility measurements in helium and nitrogen drift gas. *Int. J. Mass spectrom.* **427**, 79-90 (2018).
136. D'Atri, V. & Gabelica, V. DNA and RNA telomeric G-quadruplexes: What topology features can be inferred from ion mobility mass spectrometry? *Analyst* **144**, 6074-6088 (2019).
137. Hernandez-Alba, O., Wagner-Rousset, E., Beck, A. & Cianférani, S. Native Mass Spectrometry, Ion Mobility, and Collision-Induced Unfolding for Conformational Characterization of IgG4 Monoclonal Antibodies. *Anal. Chem.* **90**, 8865-8872 (2018).
138. Rabuck-Gibbons, J.N., Lodge, J.M., Mapp, A.K. & Ruotolo, B.T. Collision-Induced Unfolding Reveals Unique Fingerprints for Remote Protein Interaction Sites in the KIX Regulation Domain. *J. Am. Soc. Mass Spectrom.* **30**, 94-102 (2019).
139. Watanabe, Y. et al. Signature of Antibody Domain Exchange by Native Mass Spectrometry and Collision-Induced Unfolding. *Anal. Chem.* **90**, 7325-7331 (2018).
140. Mathew, A. et al. Ion Imaging of Native Protein Complexes Using Orthogonal Time-of-Flight Mass Spectrometry and a Timepix Detector. *J. Am. Soc. Mass Spectrom.* **32**, 569-580 (2021).
141. Tito, M.A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C.V. Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid. *J. Am. Chem. Soc.* **122**, 3550-3551 (2000).
142. Makarov, A. & Denisov, E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **20**, 1486-1495 (2009).
143. A., M., Vol. U.S. Patent 9043164 (May 26, 2015).
144. Smith, R.D., Cheng, X., Brace, J.E., Hofstadler, S.A. & Anderson, G.A. Trapping, detection and reaction of very large single molecular ions by mass spectrometry. *Nature* **369**, 137-139 (1994).
145. Bruce, J.E. et al. Trapping, Detection, and Mass Measurement of Individual Ions in a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer. *J. Am. Chem. Soc.* **116**, 7839-7847 (1994).

146. Chen, R. et al. Trapping, Detection, and Mass Determination of Coliphage T4 DNA Ions by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **67**, 1159-1163 (1995).
147. Benner, W.H. A Gated Electrostatic Ion Trap To Repetitiously Measure the Charge andm/zof Large Electrospray Ions. *Anal. Chem.* **69**, 4162-4168 (1997).
148. Todd, A.R., Barnes, L.F., Young, K., Zlotnick, A. & Jarrold, M.F. Higher Resolution Charge Detection Mass Spectrometry. *Anal. Chem.* **92**, 11357-11364 (2020).
149. Kafader, J.O. et al. Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nat. Methods* **17**, 391-394 (2020).
150. Mann, M., Meng, C.K. & Fenn, J.B. Interpreting Mass Spectra of Multiply Charged Ions. *Anal. Chem.* **61**, 1702-1708 (1989).
151. Ferrige, A.G. et al. Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* **6**, 707-711 (1992).
152. Reinhold, B.B. & Reinhold, V.N. Electrospray ionization mass spectrometry: Deconvolution by an Entropy-Based algorithm. *J. Am. Soc. Mass Spectrom.* **3**, 207-215 (1992).
153. Tseng, Y.-H., Uetrecht, C., Heck, A.J.R. & Peng, W.-P. Interpreting the Charge State Assignment in Electrospray Mass Spectra of Bioparticles. *Anal. Chem.* **83**, 1960-1968 (2011).
154. Zheng, H. et al. Heuristic charge assignment for deconvolution of electrospray ionization mass spectra. *Rapid Commun. Mass Spectrom.* **17**, 429-436 (2003).
155. Zhang, Z. & Marshall, A.G. A Universal Algorithm for Fast and Automated Charge State Deconvolution of Electrospray Mass-to-Charge Ratio Spectra. *J. Am. Soc. Mass Spectrom.* **9**, 225-233 (1998).
156. Stengel, F. et al. Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach. *Chem. Biol.* **19**, 599-607 (2012).
157. Van Breukelen, B., Barendregt, A., Heck, A.J.R. & Van Den Heuvel, R.H.H. Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Commun. Mass Spectrom.* **20**, 2490-2496 (2006).
158. Morgner, N. & Robinson, C.V. Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies. *Anal. Chem.* **84**, 2939-2948 (2012).
159. Tseng, Y.-H. et al. Game-Theory-Based Search Engine to Automate the Mass Assignment in Complex Native Electrospray Mass Spectra. *Anal. Chem.* **85**, 11275-11283 (2013).
160. Lu, J. et al. Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes. *J. Am. Soc. Mass Spectrom.* **26**, 2141-2151 (2015).
161. Marty, M.T. et al. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370-4376 (2015).
162. Reid, D.J. et al. MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **30**, 118-127 (2019).
163. Marty, M.T. Eliminating Artifacts in Electrospray Deconvolution with a SoftMax Function. *J. Am. Soc. Mass Spectrom.* **30**, 2174-2177 (2019).
164. Marty, M.T. A Universal Score for Deconvolution of Intact Protein and Native Electrospray Mass Spectra. *Anal. Chem.* **92**, 4395-4401 (2020).
165. Campuzano, I.D.G. et al. Native and Denaturing MS Protein Deconvolution for Biopharma: Monoclonal Antibodies and Antibody–Drug Conjugates to Polydisperse Membrane Proteins and Beyond. *Anal. Chem.* **91**, 9472-9480 (2019).
166. Bern, M. et al. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* **17**, 1216-1226 (2018).
167. Cleary, S.P., Thompson, A.M. & Prell, J.S. Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits. *Anal. Chem.* **88**, 6205-6213 (2016).
168. Cleary, S.P. et al. Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics. *J. Am. Soc. Mass Spectrom.* **29**, 2067-2080 (2018).
169. Jeong, K. et al. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell Systems* **10**, 213-218.e216 (2020).
170. Toby, T.K., Fornelli, L. & Kelleher, N.L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499-519 (2016).
171. Skinner, O.S. et al. Top-down characterization of endogenous protein complexes with native proteomics. *Nat. Chem. Biol.* **14**, 36-41 (2018).
172. Park, J. et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **14**, 909-914 (2017).

173. Liu, X. et al. Deconvolution and Database Search of Complex Tandem Mass Spectra of Intact Proteins. *Mol. Cell. Proteom.* **9**, 2772-2782 (2010).
174. Carvalho, P.C. et al. YADA: a tool for taking the most out of high-resolution spectra. *Bioinformatics* **25**, 2734-2736 (2009).
175. Horn, D., Zubarev, R. & McLafferty, F. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **11** (2000).
176. Basharat, A.R., Ning, X. & Liu, X. EnvCNN: A Convolutional Neural Network Model for Evaluating Isotopic Envelopes in Top-Down Mass-Spectral Deconvolution. *Anal. Chem.* **92**, 7778-7785 (2020).
177. Sleno, L. & Volmer, D.A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **39**, 1091-1112 (2004).
178. Brodbelt, J.S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **88**, 30-51 (2016).
179. Macias, L.A., Santos, I.C. & Brodbelt, J.S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **92**, 227-251 (2020).
180. McLuckey, S.A. & Mentinova, M. Ion/neutral, ion/electron, ion/photon, and ion/ion interactions in tandem mass spectrometry: do we need them all? Are they enough? *J. Am. Soc. Mass. Spectrom.* **22**, 3-12 (2011).
181. Benesch, J.L.P. Collisional activation of protein complexes: Picking up the pieces. *J. Am. Soc. Mass. Spectrom.* **20**, 341-348 (2009).
182. Fort, K.L. et al. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **17**, 926-933 (2018).
183. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9528-9533 (2004).
184. Brodbelt, J.S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* **43**, 2757-2783 (2014).
185. Frese, C.K. et al. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **84**, 9668-9673 (2012).
186. Riley, N.M., Westphall, M.S. & Coon, J.J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **87**, 7109-7116 (2015).
187. Brodbelt, J.S., Morrison, L.J. & Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **120**, 3328-3380 (2020).
188. Zhang, H., Cui, W., Gross, M.L. & Blankenship, R.E. Native mass spectrometry of photosynthetic pigment–protein complexes. *FEBS Lett.* **587**, 1012-1020 (2013).
189. Sobott, F. & Robinson, C.V. Characterising electrosprayed biomolecules using tandem-MS—the noncovalent GroEL chaperonin assembly. *Int. J. Mass spectrom.* **236**, 25-32 (2004).
190. Benesch, J.L. & Robinson, C.V. Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr. Opin. Struct. Biol.* **16**, 245-251 (2006).
191. Freitas, M.A., Hendrickson, C.L., Marshall, A.G., Rostom, A.A. & Robinson, C.V. Competitive binding to the oligopeptide binding protein, OppA: in-trap cleanup in an fourier transform ion cyclotron resonance mass spectrometer. *J. Am. Soc. Mass. Spectrom.* **11**, 1023-1026 (2000).
192. Speir, J.P., Senko, M.W., Little, D.P., Loo, J.A. & McLafferty, F.W. High-resolution tandem mass spectra of 37-67 kDa proteins. *J. Mass Spectrom.* **30**, 39-42 (1995).
193. Gardner, M.W. & Brodbelt, J.S. Reduction of chemical noise in electrospray ionization mass spectrometry by supplemental IR activation. *J. Am. Soc. Mass. Spectrom.* **20**, 2206-2210 (2009).
194. Greisch, J.-F., van der Laarse, S.A.M. & Heck, A.J.R. Enhancing Top-Down Analysis Using Chromophore-Assisted Infrared Multiphoton Dissociation from (Phospho)peptides to Protein Assemblies. *Anal. Chem.* **92**, 15506-15516 (2020).
195. Sahin, C., Reid, D.J., Marty, M.T. & Landreh, M. Scratching the surface: native mass spectrometry of peripheral membrane protein complexes. *Biochem. Soc. Trans.* **48**, 547-558 (2020).
196. Keener, J.E., Zhang, G. & Marty, M.T. Native Mass Spectrometry of Membrane Proteins. *Anal. Chem.* (2020).
197. Freeke, J., Robinson, C.V. & Ruotolo, B.T. Residual counter ions can stabilise a large protein complex in the gas phase. *Int. J. Mass spectrom.* **298**, 91-98 (2010).
198. Benesch, J.L.P., Aquilina, J.A., Ruotolo, B.T., Sobott, F. & Robinson, C.V. Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chem. Biol.* **13**, 597-605 (2006).
199. Stiving, A.Q. et al. Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Anal. Chem.* **91**, 190-209 (2019).

200. Zhou, M. & Wysocki, V.H. Surface Induced Dissociation: Dissecting Noncovalent Protein Complexes in the Gas phase. *Acc. Chem. Res.* **47**, 1010-1018 (2014).
201. Wang, G. et al. Releasing Nonperipheral Subunits from Protein Complexes in the Gas Phase. *Anal. Chem.* **92**, 15799-15805 (2020).
202. Popa, V., Trecroce, D.A., McAllister, R.G. & Konermann, L. Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular Dynamics Model with Mobile Protons. *J. Phys. Chem. B* **120**, 5114-5124 (2016).
203. Beardsley, R.L., Jones, C.M., Galhena, A.S. & Wysocki, V.H. Noncovalent Protein Tetramers and Pentamers with "n" Charges Yield Monomers with n/4 and n/5 Charges. *Anal. Chem.* **81**, 1347-1356 (2009).
204. Wanasundara, S.N. & Thachuk, M. Toward an Improved Understanding of the Dissociation Mechanism of Gas Phase Protein Complexes. *J. Phys. Chem. B* **114**, 11646-11653 (2010).
205. Pagel, K., Hyung, S.-J., Ruotolo, B.T. & Robinson, C.V. Alternate Dissociation Pathways Identified in Charge-Reduced Protein Complex Ions. *Anal. Chem.* **82**, 5363-5372 (2010).
206. Sciuto, S.V., Liu, J. & Konermann, L. An electrostatic charge partitioning model for the dissociation of protein complexes in the gas phase. *J. Am. Soc. Mass. Spectrom.* **22**, 1679-1689 (2011).
207. Jurchen, J.C. & Williams, E.R. Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *Journal of the American Chemical Society* **125**, 2817-2826 (2003).
208. Jones, C.M. et al. Symmetrical Gas-Phase Dissociation of Noncovalent Protein Complexes via Surface Collisions. *Journal of the American Chemical Society* **128**, 15044-15045 (2006).
209. VanAernum, Z.L. et al. Surface-Induced Dissociation of Noncovalent Protein Complexes in an Extended Mass Range Orbitrap Mass Spectrometer. *Anal. Chem.* **91**, 3611-3618 (2019).
210. Snyder, D.T. et al. Design and Performance of a Second-Generation Surface-Induced Dissociation Cell for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Native Protein Complexes. *Anal. Chem.* **91**, 14049-14057 (2019).
211. Wysocki, V.H., Joyce, K.E., Jones, C.M. & Beardsley, R.L. Surface-induced dissociation of small molecules, peptides, and non-covalent protein complexes. *J. Am. Soc. Mass. Spectrom.* **19**, 190-208 (2008).
212. Zhou, M., Dagan, S. & Wysocki, V.H. Protein Subunits Released by Surface Collisions of Noncovalent Complexes: Nativelike Compact Structures Revealed by Ion Mobility Mass Spectrometry. *Angew. Chem. Int. Ed.* **51**, 4336-4339 (2012).
213. Zhou, M., Huang, C. & Wysocki, V.H. Surface-Induced Dissociation of Ion Mobility-Separated Noncovalent Complexes in a Quadrupole/Time-of-Flight Mass Spectrometer. *Anal. Chem.* **84**, 6016-6023 (2012).
214. Harvey, S.R. et al. Relative interfacial cleavage energetics of protein complexes revealed by surface collisions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8143-8148 (2019).
215. Vimer, S. et al. Comparative Structural Analysis of 20S Proteasome Ortholog Protein Complexes by Native Mass Spectrometry. *ACS Cent. Sci.* **6**, 573-588 (2020).
216. Song, Y., Nelp, M.T., Bandarian, V. & Wysocki, V.H. Refining the Structural Model of a Heterohexameric Protein Complex: Surface Induced Dissociation and Ion Mobility Provide Key Connectivity and Topology Information. *ACS Central Science* **1**, 477-487 (2015).
217. Joly, L., Antoine, R., Broyer, M., Dugourd, P. & Lemoine, J. Specific UV photodissociation of tyrosyl-containing peptides in multistage mass spectrometry. *J. Mass Spectrom.* **42**, 818-824 (2007).
218. Madsen, J.A., Kaoud, T.S., Dalby, K.N. & Brodbelt, J.S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. *Proteomics* **11**, 1329-1334 (2011).
219. Shaw, J.B. et al. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *Journal of the American Chemical Society* **135**, 12646-12651 (2013).
220. Ly, T. & Julian, R.R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **48**, 7130-7137 (2009).
221. Reilly, J.P. Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom. Rev.* **28**, 425-447 (2009).
222. McLuckey, S.A. & Goeringer, D.E. SPECIAL FEATURE: TUTORIAL Slow Heating Methods in Tandem Mass Spectrometry. *J. Mass Spectrom.* **32**, 461-474 (1997).
223. Tamara, S. et al. Symmetry of Charge Partitioning in Collisional and UV Photon-Induced Dissociation of Protein Assemblies. *J. Am. Chem. Soc.* **138**, 10860-10868 (2016).
224. Morrison, L.J. & Brodbelt, J.S. 193 nm Ultraviolet Photodissociation Mass Spectrometry of Tetrameric Protein Complexes Provides Insight into Quaternary and Secondary Protein Topology. *Journal of the American Chemical Society* **138**, 10849-10859 (2016).

225. O'Brien, J.P., Li, W., Zhang, Y. & Brodbelt, J.S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **136**, 12920-12928 (2014).
226. Sipe, S.N. & Brodbelt, J.S. Impact of charge state on 193 nm ultraviolet photodissociation of protein complexes. *Phys. Chem. Chem. Phys.* **21**, 9265-9276 (2019).
227. Mikhailov, V.A. et al. Infrared Laser Activation of Soluble and Membrane Protein Assemblies in the Gas Phase. *Anal. Chem.* **88**, 7060-7067 (2016).
228. Zhang, Y. et al. Native MS and ECD Characterization of a Fab–Antigen Complex May Facilitate Crystallization for X-ray Diffraction. *J. Am. Soc. Mass Spectrom.* **27**, 1139-1142 (2016).
229. Lermyte, F. & Sobott, F. Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* **15**, 2813-2822 (2015).
230. Tamara, S., Scheltema, R.A., Heck, A.J.R. & Leney, A.C. Phosphate Transfer in Activated Protein Complexes Reveals Interaction Sites. *Angew. Chem.* **129**, 13829-13832 (2017).
231. Zhang, J., Loo, R.R.O. & Loo, J.A. Structural Characterization of a Thrombin-Aptamer Complex by High Resolution Native Top-Down Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **28**, 1815-1822 (2017).
232. Skinner, O.S. et al. Native Electron Capture Dissociation Maps to Iron-Binding Channels in Horse Spleen Ferritin. *Anal. Chem.* **89**, 10711-10716 (2017).
233. Li, H. et al. Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* **89**, 2731-2738 (2017).
234. Ro, S.Y. et al. Native top-down mass spectrometry provides insights into the copper centers of membrane-bound methane monooxygenase. *Nat. Commun.* **10** (2019).
235. Gupta, K. et al. Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.* **13**, 1106-1120 (2018).
236. Catherman, A.D., Skinner, O.S. & Kelleher, N.L. Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683-693 (2014).
237. Shen, X. et al. Native Proteomics in Discovery Mode Using Size-Exclusion Chromatography-Capillary Zone Electrophoresis-Tandem Mass Spectrometry. *Anal. Chem.* **90**, 10095-10099 (2018).
238. Melani, R.D. et al. Mapping Proteoforms and Protein Complexes From King Cobra Venom Using Both Denaturing and Native Top-down Proteomics. *Mol. Cell. Proteom.* **15**, 2423-2434 (2016).
239. Snijder, J. et al. Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **17**, 2522-2529 (2016).
240. Greber, B.J. et al. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* **505**, 515-519 (2014).
241. Greber, B.J. et al. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283-286 (2014).
242. Samir, P. et al. Identification of Changing Ribosome Protein Compositions using Mass Spectrometry. *Proteomics* **18**, 1800217 (2018).
243. Baßler, J. et al. Identification of a 60S Preribosomal Particle that Is Closely Linked to Nuclear Export. *Mol. Cell* **8**, 517-529 (2001).
244. Genuth, N.R. & Barna, M. The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol. Cell* **71**, 364-374 (2018).
245. Ilag, L.L. et al. Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8192-8197 (2005).
246. Rostom, A.A. et al. Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5185-5190 (2000).
247. McKay, A.R., Ruotolo, B.T., Ilag, L.L. & Robinson, C.V. Mass Measurements of Increased Accuracy Resolve Heterogeneous Populations of Intact Ribosomes. *Journal of the American Chemical Society* **128**, 11433-11442 (2006).
248. Abdillahi, A.M., Lee, K.W. & McLuckey, S.A. Mass Analysis of Macro-molecular Analytes via Multiply-Charged Ion Attachment. *Anal. Chem.* **92**, 16301-16306 (2020).
249. Stephenson, J.L. & McLuckey, S.A. Simplification of Product Ion Spectra Derived from Multiply Charged Parent Ions via Ion/Ion Chemistry. *Anal. Chem.* **70**, 3533-3544 (1998).
250. Unverdorben, P. et al. Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5544-5549 (2014).
251. Huang, X., Luan, B., Wu, J. & Shi, Y. An atomic structure of the human 26S proteasome. *Nat. Struct. Mol. Biol.* **23**, 778-785 (2016).
252. Bard, J.A.M. et al. Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **87**, 697-724 (2018).

253. Loo, J.A. et al. Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J. Am. Soc. Mass Spectrom.* **16**, 998-1008 (2005).
254. Sharon, M. et al. 20S proteasomes have the potential to keep substrates in store for continual degradation. *J. Biol. Chem.* **281**, 9569-9575 (2006).
255. Ben-Nissan, G., Vimer, S., Tarnavsky, M. & Sharon, M. Structural mass spectrometry approaches to study the 20S proteasome. *Methods Enzymol.* **619**, 179-223 (2019).
256. Sharon, M., Taverner, T., Ambroggio, X.I., Deshaies, R.J. & Robinson, C.V. Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol.* **4**, e267 (2006).
257. Sonn-Segev, A. et al. Quantifying the heterogeneity of macromolecular machines by mass photometry. *Nat. Commun.* **11**, 1772 (2020).
258. Bothner, B. & Siuzdak, G. Electrospray ionization of a whole virus: analyzing mass, structure, and viability. *ChemBioChem* **5**, 258-260 (2004).
259. Fuerstenau, S.D. et al. Mass Spectrometry of an Intact Virus. *Angew. Chem. Int. Ed. Engl.* **40**, 541-544 (2001).
260. Thomas, J.J., Bothner, B., Traina, J., Benner, W.H. & Siuzdak, G. Electrospray ion mobility spectrometry of intact viruses. *Spectrosc. - Int. J.* **18**, 31-36 (2004).
261. Kaddis, C.S. et al. Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility. *J. Am. Soc. Mass Spectrom.* **18**, 1206-1216 (2007).
262. Weiss, V.U. et al. Virus-like particle size and molecular weight/mass determination applying gas-phase electrophoresis (native nES GEMMA). *Anal. Bioanal. Chem.* **411**, 5951-5962 (2019).
263. Uetrecht, C. et al. Stability and shape of hepatitis B virus capsids in vacuo. *Angew. Chem. Int. Ed. Engl.* **47**, 6247-6251 (2008).
264. Uetrecht, C. & Heck, A.J.R. Modern Biomolecular Mass Spectrometry and its Role in Studying Virus Structure, Dynamics, and Assembly. *Angew. Chem. Int. Ed.* **50**, 8248-8262 (2011).
265. Ashcroft, A.E. Mass spectrometry-based studies of virus assembly. *Curr. Opin. Virol.* **36**, 17-24 (2019).
266. Worner, T.P., Shamorkina, T.M., Snijder, J. & Heck, A.J.R. Mass Spectrometry-Based Structural Virology. *Anal. Chem.* (2020).
267. Dulfer, J., Kadek, A., Kopicki, J.D., Krichel, B. & Uetrecht, C. Structural mass spectrometry goes viral. *Adv. Virus Res.* **105**, 189-238 (2019).
268. Uetrecht, C., Barbu, I.M., Shoemaker, G.K., van Duijn, E. & Heck, A.J.R. Interrogating viral capsid assembly with ion mobility–mass spectrometry. *Nat. Chem.* **3**, 126-132 (2011).
269. Poliakov, A. et al. Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. *J. Struct. Biol.* **157**, 371-383 (2007).
270. Veesler, D. et al. Architecture of a dsDNA viral capsid in complex with its maturation protease. *Structure* **22**, 230-237 (2014).
271. Wörner, T.P. et al. Adeno-associated virus capsid assembly is divergent and stochastic. *Nat. Commun.* **12**, 1642 (2021).
272. Snijder, J. et al. Probing the biophysical interplay between a viral genome and its capsid. *Nat. Chem.* **5**, 502-509 (2013).
273. van de Waterbeemd, M. et al. Examining the Heterogeneous Genome Content of Multipartite Viruses BMV and CCMV by Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **27**, 1000-1009 (2016).
274. Brasch, M. et al. Encapsulation of phthalocyanine supramolecular stacks into virus-like particles. *J. Am. Chem. Soc.* **133**, 6878-6881 (2011).
275. Sutter, M. et al. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939-947 (2008).
276. Gabashvili, A.N. et al. Encapsulins-Bacterial Protein Nanocompartments: Structure, Properties, and Application. *Biomolecules* **10** (2020).
277. Corchero, J.L. & Cedano, J. Self-assembling, protein-based intracellular bacterial organelles: emerging vehicles for encapsulating, targeting and delivering therapeutical cargoes. *Microb. Cell Fact.* **10**, 92 (2011).
278. Sigmund, F. et al. Bacterial encapsulins as orthogonal compartments for mammalian cell engineering. *Nat. Commun.* **9**, 1990 (2018).
279. Rurup, W.F., Snijder, J., Koay, M.S.T., Heck, A.J.R. & Cornelissen, J.J.L.M. Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. *Journal of the American Chemical Society* **136**, 3828-3832 (2014).

280. Seebeck, F.P., Woycechowsky, K.J., Zhuang, W., Rabe, J.P. & Hilvert, D. A simple tagging system for protein encapsulation. *J. Am. Chem. Soc.* **128**, 4516-4517 (2006).
281. Azuma, Y., Edwardson, T.G.W. & Hilvert, D. Tailoring lumazine synthase assemblies for bionanotechnology. *Chem. Soc. Rev.* **47**, 3543-3557 (2018).
282. Sasaki, E. et al. Structure and assembly of scalable porous protein cages. *Nat. Commun.* **8**, 14663 (2017).
283. King, N.P. & Lai, Y.T. Practical approaches to designing novel protein assemblies. *Curr. Opin. Struct. Biol.* **23**, 632-638 (2013).
284. Chen, Z. et al. De novo design of protein logic gates. *Science* **368**, 78-84 (2020).
285. Walls, A.C. et al. Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367-1382 e1317 (2020).
286. Wargacki, A.J. et al. Complete and cooperative in vitro assembly of computationally designed self-assembling protein nanomaterials. *Nat. Commun.* **12** (2021).
287. Sahasrabuddhe, A. et al. Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1268-1273 (2018).
288. Boyken, S.E. et al. De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658-664 (2019).
289. Keifer, D.Z., Motwani, T., Teschke, C.M. & Jarrold, M.F. Acquiring Structural Information on Virus Particles with Charge Detection Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **27**, 1028-1036 (2016).
290. Bajic, G., Degn, S.E., Thiel, S. & Andersen, G.R. Complement activation, regulation, and molecular basis for complement-related diseases. *EMBO J.* **34**, 2735-2757 (2015).
291. Sharp, T.H. et al. Insights into IgM-mediated complement activation based on in situ structures of IgM-C1-C4b. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11900-11905 (2019).
292. Diebolder, C.A. et al. Complement Is Activated by IgG Hexamers Assembled at the Cell Surface. *Science* **343**, 1260-1263 (2014).
293. de Jong, R.N. et al. A Novel Platform for the Potentiation of Therapeutic Antibodies Based on Antigen-Dependent Formation of IgG Hexamers at the Cell Surface. *PLoS Biol.* **14**, e1002344 (2016).
294. Wang, G. et al. Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Mol. Cell* **63**, 135-145 (2016).
295. Strasser, J. et al. Unraveling the Macromolecular Pathways of IgG Oligomerization and Complement Activation on Antigenic Surfaces. *Nano Lett.* **19**, 4787-4796 (2019).
296. Scholes, G.D., Fleming, G.R., Olaya-Castro, A. & Van Grondelle, R. Lessons from nature about solar light harvesting. *Nat. Chem.* **3**, 763-774 (2011).
297. Lu, Y., Goodson, C., Blankenship, R.E. & Gross, M.L. Primary and Higher Order Structure of the Reaction Center from the Purple Phototrophic BacteriumBlastochloris viridis: A Test for Native Mass Spectrometry. *J. Proteome Res.* **17**, 1615-1623 (2018).
298. Zhang, H. et al. Native Mass Spectrometry Characterizes the Photosynthetic Reaction Center Complex from the Purple Bacterium Rhodobacter sphaeroides. *Journal of The American Society for Mass Spectrometry* **28**, 87-95 (2017).
299. Lu, Y. et al. Native Mass Spectrometry Analysis of Oligomerization States of Fluorescence Recovery Protein and Orange Carotenoid Protein: Two Proteins Involved in the Cyanobacterial Photoprotection Cycle. *Biochemistry* **56**, 160-166 (2017).
300. Lu, Y. et al. A Molecular Mechanism for Nonphotochemical Quenching in Cyanobacteria. *Biochemistry* **56**, 2812-2823 (2017).
301. Zhang, H. et al. Native mass spectrometry and ion mobility characterize the orange carotenoid protein functional domains. *Biochim. Biophys. Acta* **1857**, 734-739 (2016).
302. Jiang, J. et al. Oligomerization state and pigment binding strength of the peridinin-Chla-protein. *FEBS Lett.* **589**, 2713-2719 (2015).
303. Zhang, H. et al. Molecular Mechanism of Photoactivation and Structural Location of the Cyanobacterial Orange Carotenoid Protein. *Biochemistry* **53**, 13-19 (2014).
304. Wen, J., Zhang, H., Gross, M.L. & Blankenship, R.E. Native Electrospray Mass Spectrometry Reveals the Nature and Stoichiometry of Pigments in the FMO Photosynthetic Antenna Protein. *Biochemistry* **50**, 3502-3511 (2011).
305. Tronrud, D.E., Schmid, M.F. & Matthews, B.W. Structure and X-ray amino acid sequence of a bacteriochlorophyll a protein from Prosthecochloris aestuarii refined at 1.9 Å resolution. *J. Mol. Biol.* **188**, 443-454 (1986).

306. Fenna, R.E. & Matthews, B.W. Chlorophyll arrangement in a bacteriochlorophyll protein from Chlorobium limicola. *Nature* **258**, 573-577 (1975).
307. Ma, J. et al. Structural basis of energy transfer in Porphyridium purpureum phycobilisome. *Nature* **579**, 146-151 (2020).
308. Zhang, J. et al. Structure of phycobilisome from the red alga Griffithsia pacifica. *Nature* **551**, 57-63 (2017).
309. Leney, A.C., Tschanz, A. & Heck, A.J.R. Connecting color with assembly in the fluorescent B-phycoerythrin protein complex. *The FEBS Journal* **285**, 178-187 (2018).
310. Saluri, M. et al. Spatial variation and structural characteristics of phycobiliproteins from the red algae Furcellaria lumbricalis and Coccotylus truncatus. *Algal Research* **52**, 102058 (2020).
311. Tamara, S., Hoek, M., Scheltema, R.A., Leney, A.C. & Heck, A.J.R. A Colorful Pallet of B-Phycoerythrin Proteoforms Exposed by a Multimodal Mass Spectrometry Approach. *Chem* **5**, 1302-1317 (2019).
312. Leney, A.C. Subunit pI Can Influence Protein Complex Dissociation Characteristics. *J. Am. Soc. Mass Spectrom.* **30**, 1389-1395 (2019).
313. Kaldmäe, M., Sahin, C., Saluri, M., Marklund, E.G. & Landreh, M. A strategy for the identification of protein architectures directly from ion mobility mass spectrometry data reveals stabilizing subunit interactions in light harvesting complexes. *Protein Sci.* **28**, 1024-1030 (2019).
314. Camacho, I.S. et al. Native mass spectrometry reveals the conformational diversity of the UVR8 photoreceptor. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1116-1125 (2019).
315. Albanese, P., Tamara, S., Saracco, G., Scheltema, R.A. & Pagliano, C. How paired PSII–LHCII supercomplexes mediate the stacking of plant thylakoid membranes unveiled by structural mass-spectrometry. *Nat. Commun.* **11** (2020).
316. Konijnenberg, A., van Dyck, J.F., Kailing, L.L. & Sobott, F. Extending native mass spectrometry approaches to integral membrane proteins. *Biol. Chem.* **396**, 991-1002 (2015).
317. Bolla, J.R., Agasid, M.T., Mehmood, S. & Robinson, C.V. Membrane Protein–Lipid Interactions Probed Using Mass Spectrometry. *Annu. Rev. Biochem.* **88**, 85-111 (2019).
318. Hopper, J.T.S. et al. Detergent-free mass spectrometry of membrane protein complexes. *Nat. Methods* **10**, 1206-1208 (2013).
319. Marty, M.T., Hoi, K.K., Gault, J. & Robinson, C.V. Probing the Lipid Annular Belt by Gas-Phase Dissociation of Membrane Proteins in Nanodiscs. *Angew. Chem. Int. Ed.* **55**, 550-554 (2016).
320. Reid, D.J. et al. Engineering Nanodisc Scaffold Proteins for Native Mass Spectrometry. *Anal. Chem.* **89**, 11189-11192 (2017).
321. Keener, J.E. et al. Chemical Additives Enable Native Mass Spectrometry Measurement of Membrane Protein Oligomeric State within Intact Nanodiscs. *Journal of the American Chemical Society* **141**, 1054-1061 (2019).
322. Walker, L.R., Marzluff, E.M., Townsend, J.A., Resager, W.C. & Marty, M.T. Native Mass Spectrometry of Antimicrobial Peptides in Lipid Nanodiscs Elucidates Complex Assembly. *Anal. Chem.* **91**, 9284-9291 (2019).
323. Walker, L.R. & Marty, M.T. Revealing the Specificity of a Range of Antimicrobial Peptides in Lipid Nanodiscs by Native Mass Spectrometry. *Biochemistry* **59**, 2135-2142 (2020).
324. Chorev, D.S. et al. Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science* **362**, 829-834 (2018).
325. Chorev, D.S. et al. The use of sonicated lipid vesicles for mass spectrometry of membrane protein complexes. *Nat. Protoc.* **15**, 1690-1706 (2020).
326. Hirst, J., Kunji, E.R.S. & Walker, J.E. Comment on "Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry". *Science* **366**, eaaw9830 (2019).
327. Chorev, D.S. & Robinson, C.V. Response to Comment on "Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry". *Science* **366**, eaax3102 (2019).
328. Alley, W.R., Jr., Mann, B.F. & Novotny, M.V. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem. Rev.* **113**, 2668-2732 (2013).
329. Ruhaak, L.R., Xu, G., Li, Q., Goonatilleke, E. & Lebrilla, C.B. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem. Rev.* **118**, 7886-7930 (2018).
330. Wuhrer, M., Catalina, M.I., Deelder, A.M. & Hokke, C.H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **849**, 115-128 (2007).
331. Čaval, T., Heck, A.J.R. & Reiding, K.R. Meta-heterogeneity: evaluating and describing the diversity in glycosylation between sites on the same glycoprotein. *Mol. Cell. Proteom.* (2020).

332. Yang, Y., Wang, G., Song, T., Lebrilla, C.B. & Heck, A.J.R. Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *MAbs* **9**, 638-645 (2017).
333. Duffin, K.L., Welply, J.K., Huang, E. & Henion, J.D. Characterization of N-linked oligosaccharides by electrospray and tandem mass spectrometry. *Anal. Chem.* **64**, 1440-1448 (1992).
334. Yang, Y., Barendregt, A., Kamerling, J.P. & Heck, A.J.R. Analyzing Protein Micro-Heterogeneity in Chicken Ovalbumin by High-Resolution Native Mass Spectrometry Exposes Qualitatively and Semi-Quantitatively 59 Proteoforms. *Anal. Chem.* **85**, 12037-12045 (2013).
335. Fussl, F., Criscuolo, A., Cook, K., Scheffler, K. & Bones, J. Cracking Proteoform Complexity of Ovalbumin with Anion-Exchange Chromatography-High-Resolution Mass Spectrometry under Native Conditions. *J. Proteome Res.* **18**, 3689-3702 (2019).
336. Hayase, T. et al. Comparison of N-glycosides of fetuins from different species and human alpha 2-HS-glycoprotein. *Biochemistry* **31**, 4915-4921 (1992).
337. Windwarder, M. & Altmann, F. Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry. *J. Proteomics* **108**, 258-268 (2014).
338. Lin, Y.H., Franc, V. & Heck, A.J.R. Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* **17**, 2861-2869 (2018).
339. Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330-333 (2020).
340. Casalino, L. et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722-1734 (2020).
341. Struwe, W.B., Stuckmann, A., Behrens, A.J., Pagel, K. & Crispin, M. Global N-Glycan Site Occupancy of HIV-1 gp120 by Metabolic Engineering and High-Resolution Intact Mass Spectrometry. *ACS Chem. Biol.* **12**, 357-361 (2017).
342. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**, 1136-1145 (2018).
343. Matsumiya, S. et al. Structural comparison of fucosylated and nonfucosylated Fc fragments of human immunoglobulin G1. *J. Mol. Biol.* **368**, 767-779 (2007).
344. Hodoniczky, J., Zheng, Y.Z. & James, D.C. Control of recombinant monoclonal antibody effector functions by Fc N-glycan remodeling in vitro. *Biotechnol. Prog.* **21**, 1644-1652 (2005).
345. Alter, G., Ottenhoff, T.H.M. & Joosten, S.A. Antibody glycosylation in inflammation, disease and vaccination. *Semin. Immunol.* **39**, 102-110 (2018).
346. Hajba, L., Szekrenyes, A., Borza, B. & Guttman, A. On the glycosylation aspects of biosimilarity. *Drug Discov. Today* **23**, 616-625 (2018).
347. Schellekens, H. & Moors, E. Biosimilars or semi-similars? *Nat. Biotechnol.* **33**, 19-20 (2015).
348. Beck, A., Sanglier-Cianferani, S. & Van Dorsselaer, A. Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Anal. Chem.* **84**, 4637-4646 (2012).
349. Thompson, N.J., Rosati, S. & Heck, A.J.R. Performing native mass spectrometry analysis on therapeutic antibodies. *Methods* **65**, 11-17 (2014).
350. Tassi, M. et al. Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products. *J. Sep. Sci.* **41**, 125-144 (2018).
351. Struwe, W.B. & Robinson, C.V. Relating glycoprotein structural heterogeneity to function – insights from native mass spectrometry. *Curr. Opin. Struct. Biol.* **58**, 241-248 (2019).
352. Rosati, S., Yang, Y., Barendregt, A. & Heck, A.J.R. Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* **9**, 967-976 (2014).
353. Parsons, T.B. et al. Optimal Synthetic Glycosylation of a Therapeutic Antibody. *Angew. Chem. Int. Ed.* **55**, 2361-2367 (2016).
354. van der Schoot, J.M.S. et al. Functional diversification of hybridoma-produced antibodies by CRISPR/HDR genomic engineering. *Sci. Adv.* **5**, eaaw1822 (2019).
355. Thompson, N.J., Hendriks, L.J.A., de Kruif, J., Throsby, M. & Heck, A.J.R. Complex mixtures of antibodies generated from a single production qualitatively and quantitatively evaluated by native Orbitrap mass spectrometry. *MAbs* **6**, 197-203 (2014).
356. Vimer, S., Ben-Nissan, G. & Sharon, M. Direct characterization of overproduced proteins by native mass spectrometry. *Nat. Protoc.* **15**, 236-265 (2020).
357. Esser-Skala, W., Wohlschlager, T., Regl, C. & Huber, C.G. A Simple Strategy to Eliminate Hexosylation Bias in the Relative Quantification of N-Glycosylation in Biopharmaceuticals. *Angew. Chem. Int. Ed.* **59**, 16225-16232 (2020).

358. Skala, W., Wohlschlager, T., Senn, S., Huber, G.E. & Huber, C.G. MoFi: A Software Tool for Annotating Glycoprotein Mass Spectra by Integrating Hybrid Data from the Intact Protein and Glycopeptide Level. *Anal. Chem.* **90**, 5728-5736 (2018).
359. Brücher, D., Franc, V., Smith, S.N., Heck, A.J.R. & Plückthun, A. Malignant tissues produce divergent antibody glycosylation of relevance for cancer gene therapy effectiveness. *MAbs* **12**, 1792084 (2020).
360. Sandra, K., Vandenheede, I. & Sandra, P. Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. *J. Chromatogr. A* **1335**, 81-103 (2014).
361. Bailey, A.O. et al. Charge variant native mass spectrometry benefits mass precision and dynamic range of monoclonal antibody intact mass analysis. *MAbs* **10**, 1214-1225 (2018).
362. Phung, W. et al. Characterization of bispecific and mispaired IgGs by native charge-variant mass spectrometry. *Int. J. Mass spectrom.* **446**, 116229 (2019).
363. Füssl, F. et al. Charge Variant Analysis of Monoclonal Antibodies Using Direct Coupled pH Gradient Cation Exchange Chromatography to High-Resolution Native Mass Spectrometry. *Anal. Chem.* **90**, 4669-4676 (2018).
364. Yan, Y., Liu, A.P., Wang, S., Daly, T.J. & Li, N. Ultrasensitive Characterization of Charge Heterogeneity of Therapeutic Monoclonal Antibodies Using Strong Cation Exchange Chromatography Coupled to Native Mass Spectrometry. *Anal. Chem.* **90**, 13013-13020 (2018).
365. Ma, F., Raoufi, F., Bailly, M.A., Fayadat-Dilman, L. & Tomazela, D. Hyphenation of strong cation exchange chromatography to native mass spectrometry for high throughput online characterization of charge heterogeneity of therapeutic monoclonal antibodies. *MAbs* **12**, 1763762 (2020).
366. Yan, Y., Xing, T., Wang, S., Daly, T.J. & Li, N. Online coupling of analytical hydrophobic interaction chromatography with native mass spectrometry for the characterization of monoclonal antibodies and related products. *J. Pharmaceut. Biomed.* **186**, 113313 (2020).
367. Schachner, L. et al. Characterization of Chain Pairing Variants of Bispecific IgG Expressed in a Single Host Cell by High-Resolution Native and Denaturing Mass Spectrometry. *Anal. Chem.* **88**, 12122-12127 (2016).
368. Yan, Y., Xing, T., Wang, S., Daly, T.J. & Li, N. Coupling Mixed-Mode Size Exclusion Chromatography with Native Mass Spectrometry for Sensitive Detection and Quantitation of Homodimer Impurities in Bispecific IgG. *Anal. Chem.* **91**, 11417-11424 (2019).
369. Gomes, F.P. & Yates III, J.R. Recent trends of capillary electrophoresis-mass spectrometry in proteomics research. *Mass Spectrom. Rev.* **38**, 445-460 (2019).
370. Belov, A.M. et al. Analysis of Proteins, Protein Complexes, and Organellar Proteomes Using Sheathless Capillary Zone Electrophoresis - Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 2614-2634 (2017).
371. Belov, A.M. et al. Complementary middle-down and intact monoclonal antibody proteoform characterization by capillary zone electrophoresis – mass spectrometry. *Electrophoresis* **39**, 2069-2082 (2018).
372. Füssl, F., Trappe, A., Carillo, S., Jakes, C. & Bones, J. Comparative Elucidation of Cetuximab Heterogeneity on the Intact Protein Level by Cation Exchange Chromatography and Capillary Electrophoresis Coupled to Mass Spectrometry. *Anal. Chem.* **92**, 5431-5438 (2020).
373. Carillo, S., Jakes, C. & Bones, J. In-depth analysis of monoclonal antibodies using microfluidic capillary electrophoresis and native mass spectrometry. *J. Pharmaceut. Biomed.* **185**, 113218 (2020).
374. de Goeij, B.E. & Lambert, J.M. New developments for antibody-drug conjugate-based therapeutic approaches. *Curr. Opin. Immunol.* **40**, 14-23 (2016).
375. Valliere-Douglass, J.F., McFee, W.A. & Salas-Solano, O. Native intact mass determination of antibodies conjugated with monomethyl Auristatin E and F at interchain cysteine residues. *Anal. Chem.* **84**, 2843-2849 (2012).
376. Jones, J., Pack, L., Hunter, J.H. & Valliere-Douglass, J.F. Native size-exclusion chromatography-mass spectrometry: suitability for antibody-drug conjugate drug-to-antibody ratio quantitation across a range of chemotypes and drug-loading levels. *MAbs* **12**, 1682895 (2020).
377. Debaene, F. et al. Innovative native MS methodologies for antibody drug conjugate characterization: High resolution native MS and IM-MS for average DAR and DAR distribution assessment. *Anal. Chem.* **86**, 10674-10683 (2014).
378. Gautier, V., Boumeester, A.J., Lössl, P. & Heck, A.J.R. Lysine conjugation properties in human IgGs studied by integrating high-resolution native mass spectrometry and bottom-up proteomics. *Proteomics* **15**, 2756-2765 (2015).

379. Botzanowski, T. et al. Insights from native mass spectrometry approaches for top- and middle- level characterization of site-specific antibody-drug conjugates. *MAbs* **9**, 801-811 (2017).
380. Fukuda, M.N., Sasaki, H., Lopez, L. & Fukuda, M. Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. *Blood* **73**, 84-89 (1989).
381. Erbayraktar, S. et al. Asialoerythropoietin is a nonerythropoietic cytokine with broad neuroprotective activity in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6741-6746 (2003).
382. Bocquet, F. et al. Biosimilar versus patented erythropoietins: learning from 5 years of European and Japanese experience. *Appl. Health. Econ. Health Policy* **13**, 47-59 (2015).
383. Yang, Y. et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 13397 (2016).
384. Čaval, T., Tian, W., Yang, Z., Clausen, H. & Heck, A.J.R. Direct quality control of glycoengineered erythropoietin variants. *Nat. Commun.* **9**, 3342 (2018).
385. Yang, Z. et al. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* **33**, 842-844 (2015).
386. Wohlschlager, T. et al. Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat. Commun.* **9**, 1713 (2018).
387. Lebede, M. et al. Exploring the Chemical Space of Protein Glycosylation in Noncovalent Protein Complexes: An Expedition Along Different Structural Levels of Human Chorionic Gonadotropin Employing Mass Spectrometry. (2021).
388. Tu, C. et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res.* **9**, 4982-4991 (2010).
389. Geyer, P.E. et al. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2**, 185-195 (2016).
390. Toby, T.K. et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nat. Protoc.* **14**, 119-152 (2019).
391. Anderson, N.L. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* **56**, 177-185 (2010).
392. Doherty, M. et al. Plasma N-glycans in colorectal cancer risk. *Sci. Rep.* **8**, 8655 (2018).
393. Qiu, Y. et al. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J. Proteome Res.* **7**, 1693-1703 (2008).
394. Pavic, T. et al. N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. *J. Transl. Med.* **16**, 323 (2018).
395. Franc, V., Yang, Y. & Heck, A.J.R. Proteoform Profile Mapping of the Human Serum Complement Component C9 Revealing Unexpected New Features of N-, O-, and C-Glycosylation. *Anal. Chem.* **89**, 3483-3491 (2017).
396. Franc, V., Zhu, J. & Heck, A.J.R. Comprehensive Proteoform Characterization of Plasma Complement Component C8alphabetagamma by Hybrid Mass Spectrometry Approaches. *J. Am. Soc. Mass. Spectrom.* **29**, 1099-1110 (2018).
397. Wu, D., Struwe, W.B., Harvey, D.J., Ferguson, M.A.J. & Robinson, C.V. N-glycan microheterogeneity regulates interactions of plasma proteins. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 8763-8768 (2018).
398. Wu, D., Li, J., Struwe, W.B. & Robinson, C.V. Probing N-glycoprotein microheterogeneity by lectin affinity purification-mass spectrometry analysis. *Chem. Sci.* **10**, 5146-5155 (2019).
399. Lin, Y.H., Zhu, J., Meijer, A.B., Vojtech, F. & Heck, A.J.R. Glycoproteogenomics: a frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/alpha-2-HS-glycoprotein. *Mol. Cell. Proteom.* **18**, 1479-1490 (2019).
400. Čaval, T. et al. Glycoproteoform Profiles of Individual Patients' Plasma Alpha-1-Antichymotrypsin are Unique and Extensively Remodeled Following a Septic Episode. *Front. Immunol.* **11**, 3413 (2021).
401. Ye, B. et al. Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin. Cancer Res.* **9**, 2904-2911 (2003).
402. Kruger, A.J. et al. Haptoglobin as an early serum biomarker of virus-induced autoimmune type 1 diabetes in biobreeding diabetes resistant and LEW1.WR1 rats. *Exp. Biol. Med. (Maywood)* **235**, 1328-1337 (2010).
403. Zhu, J. et al. Differential Quantitative Determination of Site-Specific Intact N-Glycopeptides in Serum Haptoglobin between Hepatocellular Carcinoma and Cirrhosis Using LC-EThcD-MS/MS. *J. Proteome Res.* **18**, 359-371 (2019).

404. Zhang, S., Shang, S., Li, W., Qin, X. & Liu, Y. Insights on N-glycosylation of human haptoglobin and its association with cancers. *Glycobiology* **26**, 684-692 (2016).
405. Tamara, S., Franc, V. & Heck, A.J.R. A wealth of genotype-specific proteoforms fine-tunes hemoglobin scavenging by haptoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15554-15564 (2020).
406. Buehler, P.W., Humar, R. & Schaer, D.J. Haptoglobin Therapeutics and Compartmentalization of Cell-Free Hemoglobin Toxicity. *Trends Mol. Med.* **26**, 683-697 (2020).
407. Kleppe, R. et al. Phosphorylation dependence and stoichiometry of the complex formed by tyrosine hydroxylase and 14-3-3gamma. *Mol. Cell. Proteom.* **13**, 2017-2030 (2014).
408. van de Waterbeemd, M. et al. Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. *Angew. Chem. Int. Ed. Engl.* **53**, 9660-9664 (2014).
409. Lössl, P. et al. Deciphering the Interplay among Multisite Phosphorylation, Interaction Dynamics, and Conformational Transitions in a Tripartite Protein System. *ACS Cent. Sci.* **2**, 445-455 (2016).
410. Abdul Azeez, K.R. et al. Structural mechanism of synergistic activation of Aurora kinase B/C by phosphorylated INCENP. *Nat Commun* **10**, 3166 (2019).
411. Potel, C.M., Fasci, D. & Heck, A.J.R. Mix and match of the tumor metastasis suppressor Nm23 protein isoforms in vitro and in vivo. *FEBS J.* **285**, 2856-2868 (2018).
412. Nguyen, H.H., Park, J., Kang, S. & Kim, M. Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel)* **15**, 10481-10510 (2015).
413. Puiu, M. & Bala, C. SPR and SPR Imaging: Recent Trends in Developing Nanodevices for Detection and Real-Time Monitoring of Biomolecular Events. *Sensors (Basel)* **16** (2016).
414. Greenfield, N.J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527-2535 (2006).
415. Liang, Y. Applications of isothermal titration calorimetry in protein science. *Acta Biochim. Biophys. Sin.* **40**, 565-576 (2008).
416. Ganem, B., Li, Y.T. & Henion, J.D. Detection of noncovalent receptor-ligand complexes by mass spectrometry. *Journal of the American Chemical Society* **113**, 6294-6296 (1991).
417. Loo, R.R.O., Goodlett, D.R., Smith, R.D. & Loo, J.A. Observation of a noncovalent ribonuclease S-protein/S-peptide complex by electrospray ionization mass spectrometry. *J. Am. Chem. Soc.* **115**, 4391-4392 (1993).
418. Jørgensen, T.J.D., Roepstorff, P. & Heck, A.J.R. Direct Determination of Solution Binding Constants for Noncovalent Complexes between Bacterial Cell Wall Peptide Analogues and Vancomycin Group Antibiotics by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **70**, 4427-4432 (1998).
419. J. D. Jørgensen, T., Staroske, T., Roepstorff, P., H. Williams, D. & J. R. Heck, A. Subtle differences in molecular recognition between modified glycopeptide antibiotics and bacterial receptor peptides identified by electrospray ionization mass spectrometry. *J. Chem. Soc.*, 1859-1863 (1999).
420. Cheng, X. et al. Using Electrospray Ionization FTICR Mass Spectrometry To Study Competitive Binding of Inhibitors to Carbonic Anhydrase. *Journal of the American Chemical Society* **117**, 8859-8860 (1995).
421. Feng, X., Liu, B.-F., Li, J. & Liu, X. Advances in coupling microfluidic chips to mass spectrometry. *Mass Spectrom. Rev.* **34**, 535-557 (2015).
422. Zhang, S., Van Pelt, C.K. & Wilson, D.B. Quantitative Determination of Noncovalent Binding Interactions Using Automated Nanoelectrospray Mass Spectrometry. *Anal. Chem.* **75**, 3010-3018 (2003).
423. Daniel, J.M., Friess, S.D., Rajagopalan, S., Wendt, S. & Zenobi, R. Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *Int. J. Mass spectrom.* **216**, 1-27 (2002).
424. Peschke, M., Verkerk, U.H. & Kebarle, P. Features of the ESI mechanism that affect the observation of multiply charged noncovalent protein complexes and the determination of the association constant by the titration method. *J. Am. Soc. Mass. Spectrom.* **15**, 1424-1434 (2004).
425. Schermann, S.M., Simmons, D.A. & Konermann, L. Mass spectrometry-based approaches to protein–ligand interactions. *Expert Rev. Proteom.* **2**, 475-485 (2005).
426. Van Berkel, G.J., Asano, K.G. & Schnier, P.D. Electrochemical processes in a wire-in-a-capillary bulk-loaded, nano-electrospray emitter. *J. Am. Soc. Mass. Spectrom.* **12**, 853-862 (2001).
427. Van Berkel, G.J. & Kertesz, V. Using the Electrochemistry of the Electrospray Ion Source. *Anal. Chem.* **79**, 5510-5520 (2007).
428. Zhou, S., Prebyl, B.S. & Cook, K.D. Profiling pH Changes in the Electrospray Plume. *Anal. Chem.* **74**, 4885-4888 (2002).

429. Girod, M., Dagany, X., Antoine, R. & Dugourd, P. Relation between charge state distributions of peptide anions and pH changes in the electrospray plume. A mass spectrometry and optical spectroscopy investigation. *Int. J. Mass spectrom.* **308**, 41-48 (2011).
430. Konermann, L. Addressing a Common Misconception: Ammonium Acetate as Neutral pH "Buffer" for Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **28**, 1827-1835 (2017).
431. Cech, N.B. & Enke, C.G. Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides. *Anal. Chem.* **72**, 2717-2723 (2000).
432. Robinson, C.V. et al. Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein−CoA Ligand Binding and Assembly. *Journal of the American Chemical Society* **118**, 8646-8653 (1996).
433. Wigger, M., Eyler, J.R., Benner, S.A., Li, W. & Marshall, A.G. Fourier transform-ion cyclotron resonance mass spectrometric resolution, identification, and screening of non-covalent complexes of Hck Src homology 2 domain receptor and ligands from a 324-member peptide combinatorial library. *J. Am. Soc. Mass. Spectrom.* **13**, 1162-1169 (2002).
434. Xiao, H., Kaltashov, I.A. & Eyles, S.J. Indirect assessment of small hydrophobic ligand binding to a model protein using a combination of ESI MS and HDX/ESI MS. *J. Am. Soc. Mass. Spectrom.* **14**, 506-515 (2003).
435. Gabelica, V., Galic, N., Rosu, F., Houssier, C. & De Pauw, E. Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry. *J. Mass Spectrom.* **38**, 491-501 (2003).
436. Tjernberg, A. et al. Determination of Dissociation Constants for Protein−Ligand Complexes by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **76**, 4325-4331 (2004).
437. Kempen, E.C. & Brodbelt, J.S. A Method for the Determination of Binding Constants by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **72**, 5411-5416 (2000).
438. Nguyen, G.T.H. et al. Nanoscale Ion Emitters in Native Mass Spectrometry for Measuring Ligand-Protein Binding Affinities. *ACS Cent. Sci.* **5**, 308-318 (2019).
439. Susa, A.C., Xia, Z. & Williams, E.R. Native Mass Spectrometry from Common Buffers with Salts That Mimic the Extracellular Environment. *Angew. Chem. Int. Ed.* **56**, 7912-7915 (2017).
440. Hu, J. et al. Effect of Nanoemitters on Suppressing the Formation of Metal Adduct Ions in Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **89**, 1838-1845 (2017).
441. Wortmann, A., Jecklin, M.C., Touboul, D., Badertscher, M. & Zenobi, R. Binding constant determination of high-affinity protein–ligand complexes by electrospray ionization mass spectrometry and ligand competition. *J. Mass Spectrom.* **43**, 600-608 (2008).
442. Wang, W., Kitova, E.N. & Klassen, J.S. Influence of Solution and Gas Phase Processes on Protein−Carbohydrate Binding Affinities Determined by Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **75**, 4945-4955 (2003).
443. Yu, Y., Kirkup, C.E., Pi, N. & Leary, J.A. Characterization of noncovalent protein-ligand complexes and associated enzyme intermediates of GlcNAc-6-O-sulfotransferase by electrospray ionization FT-ICR mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15**, 1400-1407 (2004).
444. Mathur, S., Badertscher, M., Scott, M. & Zenobi, R. Critical evaluation of mass spectrometric measurement of dissociation constants: accuracy and cross-validation against surface plasmon resonance and circular dichroism for the calmodulin–melittin system. *Phys. Chem. Chem. Phys.* **9**, 6187-6198 (2007).
445. Shoemaker, G.K., Soya, N., Palcic, M.M. & Klassen, J.S. Temperature-dependent cooperativity in donor-acceptor substrate binding to the human blood group glycosyltransferases. *Glycobiology* **18**, 587-592 (2008).
446. Soya, N., Shoemaker, G.K., Palcic, M.M. & Klassen, J.S. Comparative study of substrate and product binding to the human ABO(H) blood group glycosyltransferases. *Glycobiology* **19**, 1224-1234 (2009).
447. Jecklin, M.C., Schauer, S., Dumelin, C.E. & Zenobi, R. Label-free determination of protein–ligand binding constants using mass spectrometry and validation using surface plasmon resonance and isothermal titration calorimetry. *J. Mol. Recognit.* **22**, 319-329 (2009).
448. Liu, L., Kitova, E.N. & Klassen, J.S. Quantifying Protein-Fatty Acid Interactions Using Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **22**, 310-318 (2011).
449. Hofstadler, S.A. & Sannes-Lowery, K.A. Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat. Rev. Drug Discov.* **5**, 585-595 (2006).
450. Deng, G. & Sanyal, G. Applications of mass spectrometry in early stages of target based drug discovery. *Journal of Pharmaceutical and Biomedical Analysis* **40**, 528-538 (2006).

451. Hannah, V.V., Atmanene, C., Zeyer, D., Dorsselaer, A.V. & Sanglier-Cianférani, S. Native MS: an 'ESI, way to support structure- and fragment-based drug discovery. *Future Med. Chem.* **2**, 35-50 (2010).
452. Pacholarz, K.J., Garlish, R.A., Taylor, R.J. & Barran, P.E. Mass spectrometry based tools to investigate protein–ligand interactions for drug discovery. *Chem. Soc. Rev.* **41**, 4335-4355 (2012).
453. Pedro, L. & Quinn, R.J. Native Mass Spectrometry in Fragment-Based Drug Discovery. *Molecules* **21** (2016).
454. Eschweiler, J.D., Kerr, R., Rabuck-Gibbons, J. & Ruotolo, B.T. Sizing Up Protein–Ligand Complexes: The Rise of Structural Mass Spectrometry Approaches in the Pharmaceutical Sciences. *Annu. Rev. Anal. Chem.* **10**, 25-44 (2017).
455. Swayze, E.E. et al. SAR by MS: A Ligand Based Technique for Drug Lead Discovery Against Structured RNA Targets. *J. Med. Chem.* **45**, 3816-3819 (2002).
456. Maple, H.J. et al. Automated Protein–Ligand Interaction Screening by Mass Spectrometry. *J. Med. Chem.* **55**, 837-851 (2012).
457. Woods, L.A. et al. Native State Mass Spectrometry, Surface Plasmon Resonance, and X-ray Crystallography Correlate Strongly as a Fragment Screening Combination. *J. Med. Chem.* **59**, 2192-2204 (2016).
458. Gavriilidou, A.F.M., Holding, F.P., Coyle, J.E. & Zenobi, R. Application of Native ESI-MS to Characterize Interactions between Compounds Derived from Fragment-Based Discovery Campaigns and Two Pharmaceutically Relevant Proteins. *SLAS Discov.* **23**, 951-959 (2018).
459. Ren, C. et al. Quantitative Determination of Protein–Ligand Affinity by Size Exclusion Chromatography Directly Coupled to High-Resolution Native Mass Spectrometry. *Anal. Chem.* **91**, 903-911 (2019).
460. Bovet, C. et al. Estrogen receptor–ligand complexes measured by chip-based nanoelectrospray mass spectrometry: An approach for the screening of endocrine disruptors. *Protein Sci.* **16**, 938-946 (2007).
461. Jecklin, M.C. et al. Affinity Classification of Kinase Inhibitors by Mass Spectrometric Methods and Validation Using Standard IC50 Measurements. *Anal. Chem.* **81**, 408-419 (2009).
462. Maple, H.J. et al. Application of the Exactive Plus EMR for automated protein–ligand screening by non-covalent mass spectrometry. *Rapid Commun. Mass Spectrom.* **28**, 1561-1568 (2014).
463. Mehmood, S. et al. Mass spectrometry captures off-target drug binding and provides mechanistic insights into the human metalloprotease ZMPSTE24. *Nat. Chem.* **8**, 1152-1158 (2016).
464. Gavriilidou, A.F.M. et al. Native Mass Spectrometry Gives Insight into the Allosteric Binding Mechanism of M2 Pyruvate Kinase to Fructose-1,6-Bisphosphate. *Biochemistry* **57**, 1685-1689 (2018).
465. Root, K. et al. Aryl bis-sulfonamides bind to the active site of a homotrimeric isoprenoid biosynthesis enzyme IspF and extract the essential divalent metal cation cofactor. *Chemical Science* **9**, 5976-5986 (2018).
466. El-Baba, T.J. et al. Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays**. *Angew. Chem. Int. Ed.* **59**, 23544-23548 (2020).
467. Gülbakan, B. et al. Native Electrospray Ionization Mass Spectrometry Reveals Multiple Facets of Aptamer–Ligand Interactions: From Mechanism to Binding Constants. *J. Am. Chem. Soc.* **140**, 7486-7497 (2018).
468. Nguyen, G.T.H. et al. Mechanism for the Binding of Netropsin to Hairpin DNA Revealed Using Nanoscale Ion Emitters in Native Mass Spectrometry. *Anal. Chem.* (2019).
469. Gülbakan, B., Barylyuk, K. & Zenobi, R. Determination of thermodynamic and kinetic properties of biomolecules by mass spectrometry. *Curr. Opin. Biotechnol.* **31**, 65-72 (2015).
470. Daneshfar, R., Kitova, E.N. & Klassen, J.S. Determination of Protein−Ligand Association Thermochemistry Using Variable-Temperature Nanoelectrospray Mass Spectrometry. *Journal of the American Chemical Society* **126**, 4786-4787 (2004).
471. Cong, X. et al. Determining Membrane Protein–Lipid Binding Thermodynamics Using Native Mass Spectrometry. *Journal of the American Chemical Society* **138**, 4346-4349 (2016).
472. Moghadamchargari, Z. et al. Intrinsic GTPase Activity of K-RAS Monitored by Native Mass Spectrometry. *Biochemistry* **58**, 3396-3405 (2019).
473. Marchand, A., Rosu, F., Zenobi, R. & Gabelica, V. Thermal Denaturation of DNA G-Quadruplexes and Their Complexes with Ligands: Thermodynamic Analysis of the Multiple States Revealed by Mass Spectrometry. *Journal of the American Chemical Society* **140**, 12553-12565 (2018).

474. Xie, Y., Zhang, J., Yin, S. & Loo, J.A. Top-Down ESI-ECD-FT-ICR Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites. *Journal of the American Chemical Society* **128**, 14432-14433 (2006).
475. Yin, S. & Loo, J.A. Elucidating the site of protein-ATP binding by top-down mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **21**, 899-907 (2010).
476. Yin, S. & Loo, J.A. Top-Down Mass Spectrometry of Supercharged Native Protein-Ligand Complexes. *Int. J. Mass spectrom.* **300**, 118-122 (2011).
477. Clarke, D.J., Murray, E., Hupp, T., Mackay, C.L. & Langridge-Smith, P.R. Mapping a noncovalent protein-peptide interface by top-down FTICR mass spectrometry using electron capture dissociation. *J. Am. Soc. Mass. Spectrom.* **22**, 1432-1440 (2011).
478. Cammarata, M.B., Thyer, R., Rosenberg, J., Ellington, A. & Brodbelt, J.S. Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Chemical Society* **137**, 9128-9135 (2015).
479. Cammarata, M. et al. Characterization of trimethoprim resistant E. coli dihydrofolate reductase mutants by mass spectrometry and inhibition by propargyl-linked antifolates. *Chemical Science* **8**, 4062-4072 (2017).
480. Fenn, J.B. Electrospray wings for molecular elephants (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **42**, 3871-3894 (2003).
481. Vimer, S., Ben-Nissan, G. & Sharon, M. Mass Spectrometry Analysis of Intact Proteins from Crude Samples. *Anal. Chem.* **92**, 12741-12749 (2020).
482. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011).
483. The smaller the better. *Nat. Methods* **5**, 457 (2008).
484. Albrecht, T. Single-Molecule Analysis with Solid-State Nanopores. *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* **12**, 371-387 (2019).
485. Moerner, W.E. Single-Molecule Spectroscopy, Imaging, and Photocontrol: Foundations for Super-Resolution Microscopy (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **54**, 8067-8093 (2015).
486. Sage, E. et al. Neutral particle mass spectrometry with nanomechanical systems. *Nat. Commun.* **6**, 6482 (2015).
487. Yip, K.M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157-161 (2020).
488. Henderson, R. From Electron Crystallography to Single Particle CryoEM (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **57**, 10804-10825 (2018).
489. Snijder, J. et al. Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state. *Science* **355**, 1181-1184 (2017).
490. Lau, A.M. & Politis, A. Integrative Mass Spectrometry-Based Approaches for Modeling Macromolecular Assemblies. *Methods Mol. Biol.* **2247**, 221-241 (2021).
491. Bond, K., Tsvetkova, I.B., Wang, J.C., Jarrold, M.F. & Dragnea, B. Virus Assembly Pathways: Straying Away but Not Too Far. *Small* **16**, e2004475 (2020).
492. Olinares, P.D.B. et al. Native Mass Spectrometry-Based Screening for Optimal Sample Preparation in Single-Particle Cryo-EM. *Structure* **29**, 186-195 e186 (2021).

CHAPTER 3

COMPARATIVE ANALYSIS OF ANTIBODIES AND HEAVILY GLYCOSYLATED MACROMOLECULAR IMMUNE COMPLEXES BY SIZE-EXCLUSION CHROMATOGRAPHY MULTI-ANGLE LIGHT SCATTERING, NATIVE CHARGE DETECTION MASS SPECTROMETRY, AND MASS PHOTOMETRY

Maurits A. den Boer[1,2], Szu-Hsueh Lai[1,2], Xiaoguang Xue[3], Muriel D. van Kampen[3], Boris Bleijlevens[3], and Albert J.R. Heck[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

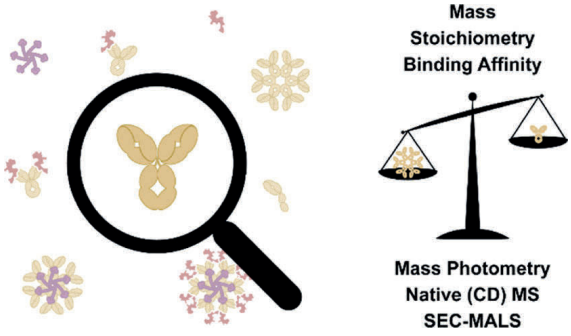
[3] Genmab, 3508 AD Utrecht, The Netherlands

Published as:

Comparative Analysis of Antibodies and Heavily Glycosylated Macromolecular Immune Complexes by Size-Exclusion Chromatography Multi-Angle Light Scattering, Native Charge Detection Mass Spectrometry, and Mass Photometry
Maurits A. den Boer, Szu-Hsueh Lai, Xiaoguang Xue, Muriel D. van Kampen, Boris Bleijlevens, and Albert J.R. Heck
Analytical Chemistry 2022 *94* (2), 892-900
DOI: 10.1021/acs.analchem.1c03656

Abstract

Qualitative and quantitative mass analysis of antibodies and related macromolecular immune complexes is a prerequisite for determining their identity, binding partners, stoichiometries, and affinities. A plethora of bioanalytical technologies exist to determine such characteristics, typically based on size, interaction with functionalized surfaces, light scattering, or direct mass measurements. While these methods are highly complementary, they also exhibit unique strengths and weaknesses. Here, we benchmark mass photometry (MP), a recently introduced technology for mass measurement, against native mass spectrometry (MS) and size exclusion chromatography multi-angle light scattering (SEC-MALS). We examine samples of variable complexity, namely, IgG4Δhinge dimerizing half-bodies, IgG-RGY hexamers, heterogeneously glycosylated IgG:sEGFR antibody-antigen complexes, and finally megadalton assemblies involved in complement activation. We thereby assess the ability to determine (1) binding affinities and stoichiometries, (2) accurate masses, for extensively glycosylated species, and (3) assembly pathways of large heterogeneous immune complexes. We find that MP provides a sensitive approach for characterizing antibodies and stable assemblies, with dissociation correction enabling us to expand the measurable affinity range. In terms of mass resolution and accuracy, native MS performs the best but is occasionally hampered by artifacts induced by electrospray ionization, and its resolving power diminishes when analyzing extensively glycosylated proteins. In the latter cases, MP performs well, but single-particle charge detection MS can also be useful in this respect, measuring masses of heterogeneous assemblies even more accurately. Both methods perform well compared to SEC-MALS, still being the most established method in biopharma. Together, our data highlight the complementarity of these approaches, each having its unique strengths and weaknesses.



Introduction

With the continued advancement of antibody-based formats as biopharmaceuticals, analytical techniques providing robust and accurate characterization of these products and related macromolecular immune complexes become increasingly important. Antibody functioning strongly depends on noncovalent protein-protein interactions, with their unique structural organization bridging molecular recognition with the recruitment of effector functions[1]. Structurally, standard IgG-based antibodies are homo-heterodimers consisting of two heavy chains (HCs) and two light chains (LCs) that are connected through several disulfide bridges. Target recognition is enabled by two variable antigen-binding (Fab) arms, which engage in highly specific epitope-paratope interactions[2,3]. Effector functions, on the other hand, are primarily mediated by the constant (Fc) tail, which recruits and directs immune cells by binding to Fc receptor proteins[4,5], but can also initiate humoral immune responses such as the classical complement pathway[6,7]. Furthermore, the Fc tail can facilitate the formation of functional oligomers – linked covalently in IgA and IgM or by noncovalent interactions in surface-bound IgGs[8]. The ability to accurately characterize antibodies and their interactions with antigens and receptors is thus of crucial importance, both for fundamental research as well as in the optimization of antibody engineering and drug development.

Affinities and kinetics of antibody-antigen interactions are typically assessed by biosensors that quantify interactions with functionalized surfaces. The most prevalent of such approaches is surface plasmon resonance[9,10], but bio-layer interferometry[11], quartz crystal microbalances[12], and Förster resonance energy transfer microscopy[13,14] provide accurate readouts down to sub-nanomolar K_d values. However, these techniques are generally limited to binary interactions, preventing their use in studying oligomerization and the formation of larger immune complexes of multiple stoichiometries and compositions.

Low-resolution biophysical methods based on size, charge, diffusion, or light scattering are widely used in academia and industry to study antibody oligomerization and complex formation. Oligomeric distributions and aggregation states can be evaluated using dynamic light scattering[15-17]. Analytical ultracentrifugation, on the other hand, can provide quantitative data at a higher resolution[18,19], allowing the technique to be used to study protein-protein interactions as well[20], but arduous experimental procedures make this approach impractical for routine use[21]. Moreover, also capillary electrophoresis-related techniques have been applied for the characterization of antibodies and their interactions[22,23].

Size exclusion chromatography (SEC) remains the long-standing industry standard for analyzing the quaternary structure of antibody products. SEC utilizes a porous matrix as a stationary phase to enable size-based separation, followed by detection to provide sensitive, highly reproducible, and quantitative data. As SEC alone does not provide an accurate means to assess masses, it often includes average molecular mass determination by coupling to a multi-angle light scattering (MALS) detector, which acts by measuring the light scattering generated by particles, being proportional to their molar mass and concentration. This makes SEC-MALS a very versatile tool for studying antibodies,

although it is also hampered by some issues because dilution and shear forces during chromatographic separation can affect equilibria. The technique also requires optimization regarding running conditions, as protein species can adsorb to the matrix.

Direct mass measurement by native mass spectrometry (MS)[24,25] represents a relatively newer component of the analytical toolbox for antibody analysis. Compared to the techniques described above, native MS yields higher mass resolution and accuracy that can be used to assess, for instance, microheterogeneity of monoclonal antibodies[26,27] and their derivatives[28-30], the formation of antibody-antigen complexes[31-33], and larger megadalton particle immune complexes[34,35]. As noncovalent interactions are retained, native MS can also be used to probe binding equilibria of protein-ligand and protein-protein interactions when instrumental parameters are carefully optimized to avoid bias and artifacts[36-40]. Direct online coupling of SEC or capillary zone electrophoresis to mass spectrometers can further enable separation and structural characterization of protein assemblies[41,42].

Typically, in native MS, mass analysis relies on resolving charge states of the same species in the *m/z* space of the mass spectrum, which becomes more difficult when the analytes become heavier and more heterogeneous. In such cases, single-particle charge-detection MS (CDMS)[43] may be very useful. This technique, which recently was also demonstrated on commercial Orbitrap-based instruments[35,44], makes it possible to directly assess the mass of single ions by measuring their *m/z* in parallel with their charge *z*, as inferred from their single-ion intensity value.

Mass photometry (MP) was recently introduced as a technology that enables mass analysis of proteins and protein complexes under native buffering conditions[45]. MP makes use of interferometric scattering microscopy to detect and quantify light scattering caused by single particles[46-50]. When particles in solution bind nonspecifically to a glass surface, their scattering signal interferes with the measured reflectivity of the glass/water interface. Because the optical properties and density of proteins are quite uniform, this reflectivity change is proportional to the molecular mass, allowing MP to provide a direct mass measurement for each particle[46,51]. MP has already been shown to be able to provide quantitative data, allowing the binding affinities and kinetics of antibody-antigen and antibody-receptor interactions to be explored[45,52,53]. However, because typical experimental conditions are limited to low nanomolar range concentrations, the methods outlined in these studies apply only to relatively strong and slowly dissociating protein assemblies.

Here, we compare MP side by side with native MS and SEC-MALS, employing these techniques for the analysis of a variety of antibody formats and heavily glycosylated macromolecular immune complexes. We expand the affinity range of quantitative MP experiments by modeling the dissociation of weaker interactions, and we compare the techniques in terms of their ability to assess a wide range of binding equilibria. We assess the pros and cons of each of these approaches, considering dynamic range, robustness, mass resolution, and mass accuracy, and we highlight their strengths in resolving extensively glycosylated species.

Materials and Methods

More detailed descriptions of the methods are provided in the **Supplementary Methods**.

Protein samples

Anti-EGFR antibodies in IgG4Δhinge, IgG1, and IgG1-RGY format and sEGFR were recombinantly expressed and purified by Genmab[34,40,54,55]. Human C1q was obtained from Complement Technology. Samples were buffer-exchanged to the appropriate solution. Protein complexes were assembled by mixing subcomponents at the desired molar ratios, followed by incubation at room temperature for at least 30 min. For quantitative experiments, incubation after preparing a dilution series was proceeded for at least 4 h.

Native MS and CDMS

For native MS, proteins in 150 mM aqueous ammonium acetate pH 7.5 were measured by direct infusion from a static nano-electrospray ionization (ESI) source. Quantitative experiments were performed on a modified LCT time-of-flight instrument (Waters), measuring samples in triplicate. All other experiments were performed on a Q Exactive Plus UHMR Orbitrap instrument (Thermo Fisher Scientific). For CDMS, dilute samples were measured at low pressure and high resolution (1 s transient) for accurate determination of both m/z and z of single ions.

SEC-MALS

SEC-MALS experiments were performed on a Waters HPLC with an in-line UV detector (Waters 2487 Dual Absorbance), a MALS detector (MiniDAWN, Wyatt Technology), and an RI detector (Optilab, Wyatt Technology). Proteins were separated on an SRT SEC-500 column (Sepax Technologies) using 100 mM sodium phosphate, 100 mM sodium sulfate, and pH 6.8 as mobile phase at 0.35 mL/min. Data were processed by ASTRA software (Wyatt) based on MALS-RI for antibody mass determination or MALS-UV-RI (Protein Conjugate Analysis) for the analysis of glycan contributions and larger complexes.

Mass Photometry

MP experiments were performed by measuring the samples in PBS on a Refeyn One^{MP} mass photometer (Refeyn). Triplicate measurements of 12,000 frames were combined into a single mass histogram. When measuring protein complexes, high concentration solutions were jump-diluted to nM range measurement concentrations in approximately 5-30 s. For quantitative experiments, a dilution series was measured in triplicate in recordings of 6000 frames. For these experiments, dissociation upon jump dilution was modeled to infer complex abundance in the original solution.

Results and Discussion

We started our MP analysis by characterizing the monomer-dimer equilibrium of hinge-deleted human IgG4 (IgG4Δhinge) molecules, providing a simple and small one-component system, for which we reported earlier data from SEC and native MS[40]. Deletion of the hinge region removes the disulfide bonds that bridge the two HCs, meaning that the two halves of the antibody interact solely via noncovalent interactions. This results in an equilibrium between antibody half molecules (HLs) and HL dimers $(HL)_2$. Previous work

from our group assessed the effects of specific mutations in the CH3 domain on this equilibrium by native MS and SEC[40], providing a panel of highly similar samples, with K_d values spanning 6 orders of magnitude (10^{-10} to 10^{-4} M). Because MP experiments are typically performed at concentrations of only a few nM, we used jump dilution to quickly dilute concentrated samples just before starting the measurement. Assuming that the k_{off} of the interaction is low enough, the observed distribution of protein assemblies should then reflect that of the original concentrated sample.

Jump-Diluted IgG4Δhinge Dimers Dissociate during MP Analysis

The distinct light scattering caused by single particles of different masses as measured by MP can be converted into masses and shown in histograms. The mass histograms of "wt" IgG4Δhinge jump-diluted from a 16 µM solution to a measurement concentration of 4 nM reveal two distinct distributions centered at the expected masses, namely, 73 kDa HL and 146 kDa $(HL)_2$ (**Figure 1A**, see **Table S1** for an overview of all measured masses). The particle counts constituted about 74% of $(HL)_2$ dimers (mass abundance of 85%), which is lower than expected for this relatively strong interaction (K_d = 50 nM by native MS[40]). Upon further inspection, we observed that the abundance of the dimer already decreased during the first seconds of the MP measurement with a pattern resembling exponential decay (**Figure 1B**). This indicates that the $(HL)_2$ dimer readily starts to dissociate upon jump dilution as the sample re-equilibrates to the measurement concentration, meaning that simply summing the data of the measurement window will lead to an underestimation of the actual dimer abundance.

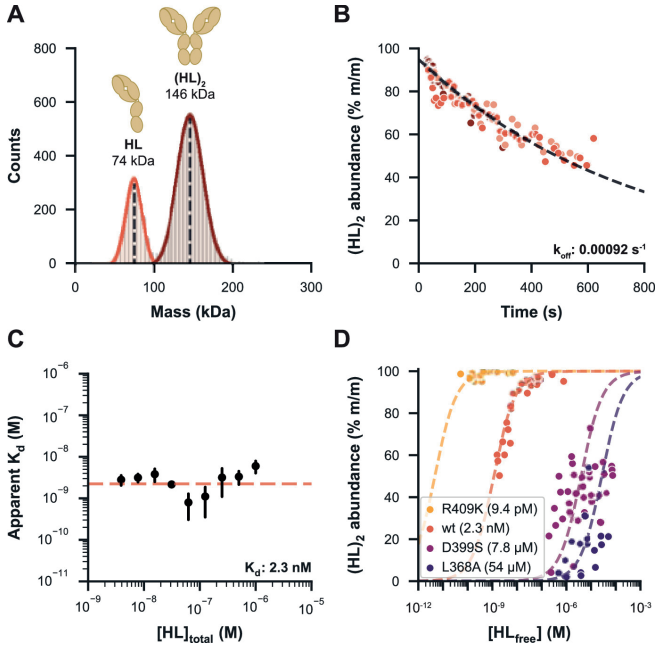


Figure 1 | Qualitative and quantitative characterization of IgG4Δhinge mutants by MP. (A) Mass histogram showing particle counts of "wt" IgG4Δhinge in PBS, jump diluted from 16 µM and measured at 4 nM, with normal distributions fitted for HL (bright red, 26%) and $(HL)_2$ (dark red, 74%). This histogram corresponds to the first 80 s after jump dilution. Masses are indicated as the mean of a fitted normal distribution. **(B)** Monomer-dimer distribution during an extended experiment in triplicate (shades of red) revealed that the abundance of "wt" $(HL)_2$ decreased during the analysis time window. Data were split into bins of 100 events, and an exponential decay function was fitted to the dimer abundance within the bin to determine the k_{off}. **(C).** Determined k_{off} was used to estimate the ratio between HL and $(HL)_2$ at the instant of jump dilution for a dilution series of the "wt" measured in triplicate, revealing the apparent K_d of each measurement, followed by the calculation of a K_d value for the whole dilution series. **(D).** Fractional dimer abundances and K_d values resulting from a dilution series of four IgG4Δhinge mutational variants, demonstrating that MP can assess affinities over a broad dynamic range.

Modelling for Dissociation Expands the Affinity Range of IgG4Δhinge Mutants Assessable by MP

To obtain a more accurate representation of the oligomer distributions in solution before jump dilution, we adjusted our data processing approach by modeling $(HL)_2$ dissociation. When the sample concentration is diluted by several orders of magnitude, especially when [HL] << K_d, we can assume that the initial decrease in $(HL)_2$ abundance is driven primarily by k_{off}. Thus, by determining k_{off} of the interaction, we can fit an exponential decay function to the measured $[(HL)_2]$ over time to estimate $[(HL)_2]$ before jump dilution (see **Supplementary Methods**). We evaluated this method using a panel of four IgG4Δhinge mutants spanning a broad affinity range, measuring a dilution series to determine their K_d (**Figure 1C, D**). Modelling for dissociation allowed us to determine affinities well into the µM range, substantially improving the dynamic range of quantitative MP experiments. Still, a few issues remained. We found that the most consistent results were obtained for relatively strong interactions with low k_{off} values, such as those of "wt" IgG4Δhinge (K_d =

2.3 nM). For the even stronger interactions of the R409K mutant (K_d = 9.4 pM), the equilibrium was still mostly geared toward the dimer at concentrations assessable by MP, reducing the precision of K_d determination for this mutant. Weaker interactions such as those of the D399S (K_d = 7.8 μM) and L368A (K_d = 54 μM) mutants could also be measured, although higher dissociation rates (0.012 and 0.029 s^{-1}) reduced the accuracy of the model. Nonetheless, MP experiments led to the same affinity ranking of the mutants as did native MS, although with some discrepancies between the obtained absolute K_d values (**Table S2**). An important difference is that native MS is performed with a volatile buffering solution (*i.e.*, aqueous ammonium acetate), while MP enabled the use of PBS as a more physiological buffer. Furthermore, standard native MS is also somewhat restrained to high nM to low μM concentrations, reducing the accuracy by which (sub) nM range K_d values can be assessed. MP and native MS are highly complementary in this sense, as they each have their own distinctive preferred concentration range.

MP Outperforms Native MS in the Mass Assessment of Heavily Glycosylated Antibody-Antigen Assemblies

We next characterized the interactions between antibodies and their antigen by MP and native MS. Several therapeutic antibodies target glycosylated receptor proteins, some of which are notoriously hard to analyze by native MS due to their high degree of microheterogeneity. Methods that can accurately mass measure and quantify these antigens and their interaction with mAbs are therefore of great use to both fundamental and biopharmaceutical research. Here, we analyze an IgG1 mAb targeting epidermal growth factor receptor (EGFR), whereby we used the soluble domain (sEGFR). This protein with a mass of 69,409 Da in its non-glycosylated form is very heterogeneous, harboring 11 typical *N*-glycosylation motifs that can be variably occupied[56].

Starting with the IgG1 alone, although the average mass obtained by MP was in good agreement with native MS, the latter provided unparalleled mass accuracy and resolution, enabling baseline resolution of individual glycoforms (**Figure 2**). However, native mass analysis of extensively glycosylated sEGFR alone was difficult, being unable to resolve charge states because of the presence of a plethora of proteoglycoforms. This obstacle was overcome by using charge-detection MS (CDMS), recently developed as a methodology for Orbitrap instruments, which provides an extra dimension of data by measuring the charge of the ions independently. CDMS measured a mass of 88 kDa for sEGFR, in close agreement with an earlier reported value derived by tandem MS experiments[30]. Similarly, MP readily provided a mass of 86 kDa, with SEC-MALS-UV-RI also measuring a mass of 91 kDa for sEGFR (**Figure S3**). Next, when the anti-sEGFR mAb was incubated together with sEGFR, MP presented further advantages. In native MS, additional ion signals were observed for $(IgG1)_1$:$(sEGFR)_1$ (*m/z* 7000-8500), although poor resolution hampered mass determination, while the full $(IgG1)_1$:$(sEGFR)_2$ complex (*m/z* 8500-10,000) could not be resolved at all. However, both CDMS and jump dilution MP enabled the reliable measurement of the average masses for all co-occurring complexes, clearly revealing the stoichiometry. However, binding occupancy was somewhat lower in CDMS, likely because of re-equilibration upon dilution before the somewhat longer measurements. SEC-MALS

was similarly able to discern the full $(IgG1)_1$:$(sEGFR)_2$ complex, although the resolution was substantially lower (**Figure S3**). Combining them, these data already show that MP and CDMS have advantages for mass analysis of heterogeneous antibody-antigen complexes.

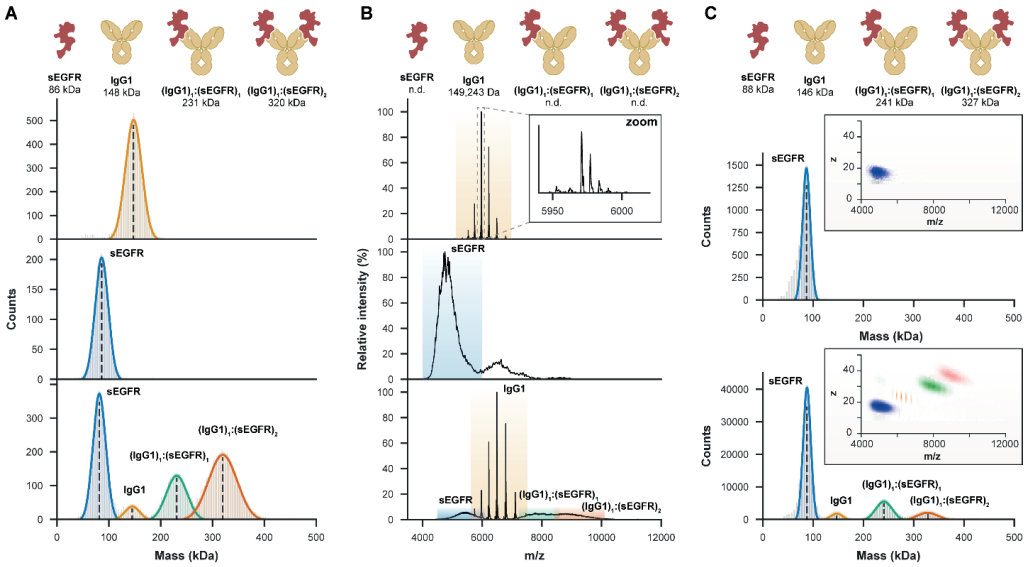


Figure 2 | MP and CDMS may overcome certain limitations of native MS in the mass measurements of highly heterogeneous antibody-antigen complexes. (A) MP provides an average mass for IgG1 (upper panel) and sEGFR (middle) and is not hampered by the high micro-heterogeneity of the latter. When 2 µM IgG1 was incubated with 5 µM of sEGFR to form $(IgG1)_1$:$(sEGFR)_1$ and $(IgG1)_1$:$(sEGFR)_2$ complexes, jump dilution MP could resolve these highly heterogeneous species (lower). **(B)** Although native MS on samples at the same concentrations provided superior mass resolution and accuracy for free IgG1 (upper), resolving individual glycoforms (zoom), the high microheterogeneity of sEGFR, measured separately (middle) and in antibody-antigen complexes (lower), resulted in unresolved features. In these experiments, overlapping charge states prevented mass measurements of these species. **(C)** More accurate masses could be obtained by native CDMS, measuring in two dimensions *m/z* and *z* (insets) for sEGFR (upper) and all co-occurring species involving IgG1 and sEGFR (lower). For these experiments, the same native MS samples were diluted 20-fold, leading to re-equilibration and thus a lower binding occupancy.

MP, SEC, and Native MS Analyses of the Monomer -Hexamer Equilibrium of Soluble IgG1-RGY Hexamers Produce Consistent Results

We next evaluated the performance of MP in the characterization of larger and more complex antibody-based systems, involved in immune activation through the complement pathway[8,34]. Target-bound IgG can initiate complement activity by forming a hexameric binding platform for recognition of complex C1q. Although these IgG oligomers are thought to only form by clustering on antigenic surfaces *in vivo*, this process can be mimicked in solution using the engineered IgG-RGY platform, a triple mutant that readily forms hexamers in equilibrium with monomers[8,34,55].

MP mass histograms and SEC-MALS chromatograms of IgG1-RGY revealed as expected two species corresponding to the monomer (denoted $(IgG1)_1$) and hexamer ($(IgG1)_6$)

(**Figure 3A, B**). While the MP mass of the monomer was in good agreement with the native MS data (**Figure 3C**), we noticed that the mass of the hexamer was consistently off by about + 70 kDa, possibly due to its non-globular shape as a flat disc. In SEC-MALS, we observed peak trailing for the hexamer, potentially driven by shear force-induced dissociation or re-equilibration during separation. As reported previously[8,34], native mass spectra of IgG1-RGY also showed two well-resolved distributions for $(IgG1)_1$ and $(IgG1)_6$, but uniquely also intermediate oligomers at lower abundance. Possibly the ESI process could result in partial dissociation of the hexamers, as this process is distinct from gas-phase-based tandem MS (**Figure S4**).

Quantifying the abundance of the hexamer in a dilution series of IgG1-RGY by all three techniques resulted in highly comparable data, although with some subtle differences. In agreement with earlier studies[57], longer MP recordings showed that hexamers re-equilibrate only very slowly upon jump dilution (**Figure S5**), meaning that such MP experiments should directly provide an accurate representation of the monomer-hexamer equilibrium. To characterize the equilibrium of IgG1-RGY by MP, we measured a dilution series and compared results to SEC and native MS. MP measurements proved to be quite consistent between replicates and could be performed down to nM range concentrations that cannot be assessed by native MS or SEC (**Figure 3A**). While SEC also proved to provide very robust data, hexamer abundances were fractionally lower, potentially due to dissociation during separation. Finally, although native MS performed well at higher concentrations, variability increased at lower concentrations. Nonetheless, each of the three techniques revealed that about half of the IgGs are incorporated into hexamers at a concentration of 1 µM, consistent with previously reported data[34].

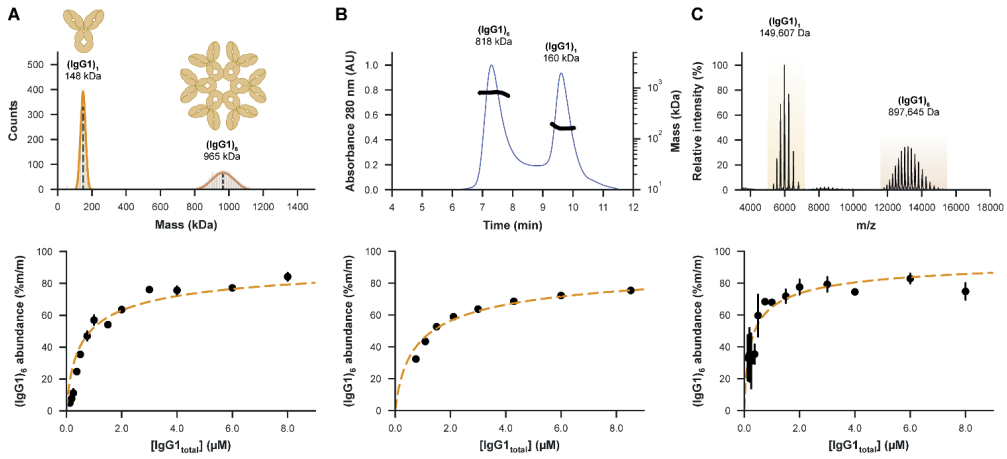


Figure 3 | MP enables qualitative and quantitative characterization of the monomer-hexamer equilibrium of IgG1-RGY. (A) MP mass histogram (top) of 2 µM IgG1-RGY in PBS jump diluted to 10 nM showing monomeric ($(IgG1)_1$) and hexameric ($(IgG1)_6$) species. The mass of the hexamer was consistently measured about 70 kDa too high. The relative abundance of the IgG1-RGY hexamer was measured over a dilution series spanning a concentration range of 0.1 to 8 µM (bottom), with error bars indicating the standard deviation over three technical replicate measurements. **(B)** SEC-MALS chromatogram of the same 2 µM IgG1-RGY sample (top) and the fractional abundance of the hexamer as measured by SEC-UV over a dilution series (bottom), revealing an alike monomer to hexamer ratio. The quantitative data in the lower panel is adapted from van Kampen *et al.*[58]. **(C)** Native mass spectrum (top) of 2 µM IgG1-RGY measured in 150 mM

NH_4OAc pH 7.5, revealing two distinct ion series for the monomer and hexamer, with ions originating from intermediate oligomeric states observed at lower abundance. While generally in good agreement with the other methods, hexamer abundances measured by MS (bottom) were less consistent and higher than expected, particularly at the lowest measured concentrations.

Characterization of Complement Component C1q by MP Exposes Shortcomings of Native MS and SEC-MALS

We next characterized complement component C1q, the recognition complex of the classical complement pathway, revealing striking differences between the three tested techniques. C1q is a 464 kDa 18-membered protein complex that consists of three pairs of triple helices ($A_2B_2C_2$) that are joined in a stem and end with six globular headpieces[59,60]. Although each headpiece has a low affinity for the Fc of IgGs[61,62], clustering into oligomers allows for multivalent binding with increased avidity[63], making the six-armed C1q complex highly compatible with IgM[64] and IgG hexamers[8]. As demonstrated previously, C1q behaves anomalously in SEC, eluting at a short retention time that suggests a mass of > 1 MDa[34] (**Figure S6A**). Because of its open structure, C1q may have a much larger hydrodynamic radius than globular proteins of similar mass, producing a bias in size-based separation. Concordantly, the coupled MALS system revealed that this elution peak did correspond to the free C1q complex with a mass of about 444 kDa. Curiously, the native mass spectra of C1q were consistently marked by the presence of three main ion series corresponding to two-armed, four-armed, and complete six-armed C1q species (**Figure S6B**), decreasing in abundance with the size of the complex. Partial C1q complexes did not display asymmetrical charge partitioning characteristic for collisional dissociation in the gas phase, suggesting that, similar to intermediate oligomers of IgG-RGY, smaller complexes are formed by the electrospray process. Lastly, MP measurements of C1q revealed predominantly particles of the intact six-armed C1q complex ($A_6B_6C_6$), with a minor contribution being made by a two-armed $A_2B_2C_2$ complex (**Figure S6C**). We thus conclude that MP seems the most unbiased tool for the analysis of C1q.

MP and CDMS Tackle Mass Heterogeneity When Analyzing Immune Complexes of IgG-RGY Hexamers Bound to Highly Glycosylated Antigens and C1q

Having demonstrated that IgG1-RGY hexamers and C1q can be measured accurately by MP, we next sought to characterize immune activation complexes involving antigen-bound IgG1-RGY hexamers and C1q. When IgG1-RGY was first incubated with C1q, MP revealed the formation of a 1.43 MDa complex corresponding to $(IgG1)_6$:$(C1q)_1$ (**Figure 4A**). Similar to measurements of the IgG1-RGY hexamers, this mass is about 60 kDa higher than expected, possibly due to the non-globular shapes of both complexes. Next, we assembled larger complement activation complexes associated with highly heterogeneous sEGFR antigens, whereby sEGFR was incubated with preformed $(IgG1)_6$:$(C1q)_1$ complexes (see **Figure S7** for the analysis of IgG1-RGY with sEGFR separately). MP revealed the presence of particles with an average mass of 2.35 MDa, likely corresponding to $(sEGFR)_{12}$:$(IgG1)_6$:$(C1q)_1$. However, as the resulting peak was quite broad, we could not yet exclude the possibility of other (co-occurring) stoichiometries. Similar to MP, SEC-MALS-UV-RI measurement of IgG1-RGY incubated with C1q led to the detection of

$(IgG1)_6$:$(C1q)_1$ with a fairly adequate mass of 1.3 MDa (**Figure 4B**). Although larger complexes of ~ 1.9 MDa were observed upon the addition of sEGFR, here, the width of this peak and the accuracy of the mass measurement was insufficient for determining the exact stoichiometry, possibly also due to dissociation of sEGFR during separation. In agreement with earlier reports[34], native MS of IgG1-RGY with C1q alone revealed the presence of $(IgG1)_6$:$(C1q)_1$ complexes, for which an accurate mass of 1361 kDa could be established (**Figure 4C**). Contrary to MP, however, relatively more $(IgG1)_6$ and uniquely also $(IgG1)_6$:$(A_4B_4C_4)_1$ complexes were observed, potentially dissociation products of the full complex formed in the MS source region. While larger but unresolved ions could be detected by native MS after the addition of sEGFR, an inability to resolve charge states prevented mass determination. For these samples, CDMS enabled confident assignment of a mass of 2.42 MDa, corresponding to the complete $(sEGFR)_{12}$:$(IgG1)_6$:$(C1q)_1$ assembly (**Figure 4D**).

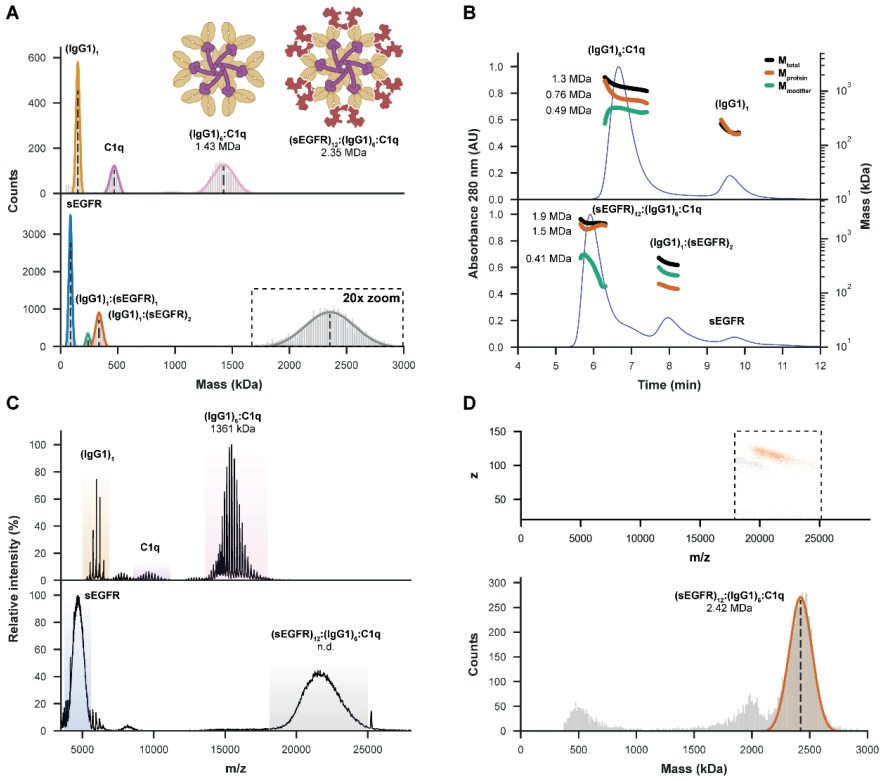


Figure 4 | MP and CDMS successfully determine the mass and stoichiometry of highly heterogeneous $(sEGFR)_{12}$:$(IgG1)_6$:C1q immune complexes. (A) MP measurements of IgG1-RGY incubated with C1q reveal the formation of $(IgG1)_6$:$(C1q)_1$ complexes, with nearly all IgG hexamers occupied. When incubating C1q with pre-formed $(IgG1)_6$:$(sEGFR)_{12}$, MP resolves a 2.35 MDa complex, likely corresponding to $(sEGFR)_{12}$:$(IgG1)_6$:$(C1q)_1$. **(B)** SEC-MALS-UV-RI analysis similarly reveals the formation of ~ 1.3 MDa $(IgG1)_6$:$(C1q)_1$ (with $(IgG1)_6$ measured as the 0.76 MDa "protein" and C1q as a 0.49 MDa "modifier"). When sEGFR was added, SEC-MALS-UV-RI revealed the formation of larger complexes of around 1.9 MDa (1.5 MDa for $(sEGFR)_{12}$:$(IgG1)_6$ with a 0.41 MDa modifier). **(C)** Measurement of the same samples by native MS reveals an accurate mass for $(IgG1)_6$:$(C1q)_1$, but the technique struggles with complexes involving sEGFR. Larger ion species were detected in such experiments, but they could not be charge-resolved. **(D)** Single-particle measurements of the distribution around *m/z* 21,000 by CDMS (top) revealed a mass of 2.42 MDa (bottom) corresponding to the expected mass of the full $(sEGFR)_{12}$:$(IgG1)_6$:$(sEGFR)_{12}$ complex (bottom).

Comparing Advantages and Disadvantages Reveals That MP, Native MS, and SEC-MALS Are Highly Complementary

Reflecting on the analyses performed in this study on a wide variety of systems, we can compare the advantages and disadvantages of the approaches (see **Table S1** for an overview of all the measured masses and **Table S3** for a qualitative comparison between the techniques). Analytical SEC-MALS is the most established method, providing reproducible, robust, and quantitative measurements. While accurate for smaller proteins, masses of larger multicomponent systems were underestimated, often by as much as 10-20%, making SEC-MALS suboptimal for large complexes that may dissociate by shear stress or dilution effects during column separation. MP represents a relatively new

approach that is a fast and comparatively straightforward technique to measure more accurate masses in native-like buffering solutions, enabling it to tackle multicomponent systems more effectively than SEC-MALS. Still, mass resolution limits the technology mostly to antibody-protein interactions, as mass differences induced by small molecule or peptide binding are mostly too small to resolve. Sample consumption is low, however, and unlike conventional native MS, the technique is not hampered by extensive protein glycosylation. While MP could reliably measure molecular masses and quantify strongly interacting and slowly dissociating protein complexes, jump dilution to nM range concentrations induced dissociation of weaker interactions. Another consideration for using MP is that mass measurements may be affected by the shape of the particles, as we consistently measured a higher-than-expected mass for IgG1-RGY hexamers. Native MS provided superior mass resolution and accuracy, resolving individual proteoglycoforms in samples of moderate complexity and uniquely providing unambiguous stoichiometries for protein complexes. After careful optimization of instrumental parameters, reliable quantification could be achieved when measuring at conventionally used concentrations (around 10^{-6} M). Certain protein complexes, however, proved to be sensitive to dissociation induced in solution by the ESI process, and performance of conventional native MS deteriorated when analyzing extensively glycosylated proteins. However, when charge states could not be resolved in the *m/z* domain, this challenge was effectively overcome by CDMS, providing unparalleled mass accuracy for highly heterogeneous protein assemblies.

Conclusions

Here, we compared MP with native MS and SEC-MALS for the qualitative and quantitative analysis of antibodies and related immune complexes. Single-molecule and solution-based MP provides a relatively straightforward way to assess protein complexes and can fill gaps between the two other techniques in terms of mass accuracy and resolution, while also being able to quantitatively assess strong and stable protein-protein interactions. Among the main benefits of MP are its high sensitivity (nM) and ability to measure in-solution using a wide variety of buffer solutions. However, here we show that the mass resolving power of MP is still somewhat limited and that some protein complexes may dissociate due to dilution required for measurements. Among the main benefits of native MS are its comparatively very high mass resolution, but only when charge states can be resolved. When high heterogeneity makes this impossible, single-molecule CDMS can be used to infer charge states in an alternative manner, providing lower resolving power than conventional native MS, but generally still higher than MP. Native MS is, however, a gas-phase technique that requires volatile buffers and has a low tolerance for salts and detergents. SEC and SEC-MALS are well-established technologies for mass assessment of proteins within the biopharmaceutical laboratories, benefiting from their proven robustness and ease of use. However, for assessing protein assemblies, the resolving power of SEC-MALS is relatively low, and quantification may be somewhat hampered by dilution of the sample during separation. Overall, our data show that the tested approaches are highly complementary, each having its unique preferred use cases. With

robust commercial instruments now becoming available, these newer techniques may become more accepted. Furthermore, combining techniques, such as SEC coupled to native MS, may overcome some of their weaknesses, while taking advantage of their strengths.

Supporting Information

Contains supplementary methods, k_{off} and K_d determination for IgG4Δhinge variants, SEC-MALS of IgG1 and sEGFR, native MS gas-phase dissociation of IgG1 hexamers, IgG1 hexamer dissociation upon jump dilution MP, SEC-MALS, native MS and MP characterization of C1q, SEC-MALS, native (CD) MS and MP characterization of $(IgG1)_6$:$(sEGFR)_{12}$ antibody-antigen complexes, comparison of masses measured by the assessed techniques, MP-derived kinetic rates and K_d values for IgG4Δhinge variants, comparison of advantages and disadvantages of the assessed techniques and supplementary references.

Conflicts of interest

XX, MDvK and BB are Genmab employees and own Genmab warrants and/or stock.

Acknowledgements

We especially thank Genmab research associates Mandy Blom and Clifford Rodriguez for their excellent work on performing the SEC-MALS experiments and fruitful discussions. We further thank members of the Heck laboratory for general support, especially Arjan Barendregt. This research received funding through the Netherlands Organization for Scientific Research (NWO) TTW-NACTAR project 16442 (AJRH and MAdB) and the Spinoza Award SPI.2017.028 to AJRH. We further acknowledge funding for the large-scale proteomics facility, the Netherlands Proteomics Center, through the X-omics Road Map program (Project 184.034.019) and the EU Horizon 2020 program Epic-XS (Project 823839).

References

1. Schroeder, H.W., Jr. & Cavacini, L. Structure and function of immunoglobulins. *J Allergy Clin Immunol* **125**, S41-52 (2010).
2. Peng, H.P., Lee, K.H., Jian, J.W. & Yang, A.S. Origins of specificity and affinity in antibody-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2656-2665 (2014).
3. Sela-Culang, I., Kunik, V. & Ofran, Y. The Structural Basis of Antibody-Antigen Recognition. *Front. Immunol.* **4**, 302 (2013).
4. Hayes, J.M., Wormald, M.R., Rudd, P.M. & Davey, G.P. Fc gamma receptors: glycobiology and therapeutic prospects. *J Inflamm Res* **9**, 209-219 (2016).
5. Mellor, J.D., Brown, M.P., Irving, H.R., Zalcberg, J.R. & Dobrovic, A. A critical review of the role of Fc gamma receptor polymorphisms in the response to monoclonal antibodies in cancer. *J. Hematol. Oncol.* **6**, 1 (2013).
6. Ugurlar, D. et al. Structures of C1-IgG1 provide insights into how danger pattern recognition activates complement. *Science* **359**, 794-797 (2018).
7. Lu, L.L., Suscovich, T.J., Fortune, S.M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nat. Rev. Immunol.* **18**, 46-61 (2018).
8. Diebolder, C.A. et al. Complement is activated by IgG hexamers assembled at the cell surface. *Science* **343**, 1260-1263 (2014).
9. Nguyen, H.H., Park, J., Kang, S. & Kim, M. Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel)* **15**, 10481-10510 (2015).
10. Puiu, M. & Bala, C. SPR and SPR Imaging: Recent Trends in Developing Nanodevices for Detection and Real-Time Monitoring of Biomolecular Events. *Sensors (Basel)* **16**, 870 (2016).
11. Kamat, V. & Rafique, A. Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. *Anal. Biochem.* **536**, 16-31 (2017).
12. Ogi, H. Wireless-electrodeless quartz-crystal-microbalance biosensors for studying interactions among biomolecules: a review. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **89**, 401-417 (2013).
13. Goldschen-Ohm, M.P., White, D.S., Klenchin, V.A., Chanda, B. & Goldsmith, R.H. Observing Single-Molecule Dynamics at Millimolar Concentrations. *Angew Chem Int Ed Engl* **56**, 2399-2402 (2017).
14. Aggarwal, V. & Ha, T. Single-molecule fluorescence microscopy of native macromolecular complexes. *Curr. Opin. Struct. Biol.* **41**, 225-232 (2016).
15. Nobbmann, U. et al. Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnol. Genet. Eng. Rev.* **24**, 117-128 (2007).
16. Stetefeld, J., McKenna, S.A. & Patel, T.R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409-427 (2016).
17. Hanlon, A.D., Larkin, M.I. & Reddick, R.M. Free-Solution, Label-Free Protein-Protein Interactions Characterized by Dynamic Light Scattering. *Biophys. J.* **98**, 297-304 (2010).
18. Berkowitz, S.A. & Philo, J.S. in Biophysical Characterization of Proteins in Developing Biopharmaceuticals. (eds. D.J. Houde & S.A. Berkowitz) 211-260 (Elsevier, Amsterdam; 2015).
19. Uchiyama, S., Noda, M. & Krayukhina, E. Sedimentation velocity analytical ultracentrifugation for characterization of therapeutic antibodies. *Biophys. Rev.* **10**, 259-269 (2018).
20. Schuck, P. Analytical Ultracentrifugation as a Tool for Studying Protein Interactions. *Biophys. Rev.* **5**, 159-171 (2013).
21. Gandhi, A.V., Pothecary, M.R., Bain, D.L. & Carpenter, J.F. Some Lessons Learned From a Comparison Between Sedimentation Velocity Analytical Ultracentrifugation and Size Exclusion Chromatography to Characterize and Quantify Protein Aggregates. *J. Pharm. Sci.* **106**, 2178-2186 (2017).
22. Moser, A.C., Trenhaile, S. & Frankenberg, K. Studies of antibody-antigen interactions by capillary electrophoresis: A review. *Methods* **146**, 66-75 (2018).
23. Kumar, R., Guttman, A. & Rathore, A.S. Applications of capillary electrophoresis for biopharmaceutical product characterization. *Electrophoresis* (2021).
24. Leney, A.C. & Heck, A.J. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass. Spectrom.* **28**, 5-13 (2017).
25. Tamara, S., den Boer, M.A. & Heck, A.J.R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* (2021).
26. Thompson, N.J., Hendriks, L.J.A., de Kruif, J., Throsby, M. & Heck, A.J.R. Complex mixtures of antibodies generated from a single production qualitatively and quantitatively evaluated by native Orbitrap mass spectrometry. *MAbs* **6**, 197-203 (2014).

27. Yang, Y. et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat Commun* **7**, 13397 (2016).
28. Valliere-Douglass, J.F., McFee, W.A. & Salas-Solano, O. Native Intact Mass Determination of Antibodies Conjugated with Monomethyl Auristatin E and F at Interchain Cysteine Residues. *Anal. Chem.* **84**, 2843-2849 (2012).
29. Hengel, S.M. et al. Measurement of in Vivo Drug Load Distribution of Cysteine-Linked Antibody–Drug Conjugates Using Microscale Liquid Chromatography Mass Spectrometry. *Anal. Chem.* **86**, 3420-3425 (2014).
30. Dyachenko, A. et al. Tandem Native Mass-Spectrometry on Antibody-Drug Conjugates and Submillion Da Antibody-Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **87**, 6095-6102 (2015).
31. Atmanene, C. et al. Extending Mass Spectrometry Contribution to Therapeutic Monoclonal Antibody Lead Optimization: Characterization of Immune Complexes Using Noncovalent ESI-MS. *Anal. Chem.* **81**, 6364-6373 (2009).
32. Debaene, F. et al. Time Resolved Native Ion-Mobility Mass Spectrometry to Monitor Dynamics of IgG4 Fab Arm Exchange and "Bispecific" Monoclonal Antibody Formation. *Anal. Chem.* **85**, 9785-9792 (2013).
33. Wang, G., de Jong, R.N., van den Bremer, E.T.J., Parren, P.W.H.I. & Heck, A.J.R. Enhancing Accuracy in Molecular Weight Determination of Highly Heterogeneously Glycosylated Proteins by Native Tandem Mass Spectrometry. *Anal. Chem.* **89**, 4793-4797 (2017).
34. Wang, G. et al. Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Mol. Cell* **63**, 135-145 (2016).
35. Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **17**, 395-398 (2020).
36. Eschweiler, J.D., Kerr, R., Rabuck-Gibbons, J. & Ruotolo, B.T. Sizing Up Protein–Ligand Complexes: The Rise of Structural Mass Spectrometry Approaches in the Pharmaceutical Sciences. *Annu. Rev. Anal. Chem.* **10**, 25-44 (2017).
37. Mathur, S., Badertscher, M., Scott, M. & Zenobi, R. Critical evaluation of mass spectrometric measurement of dissociation constants: accuracy and cross-validation against surface plasmon resonance and circular dichroism for the calmodulin–melittin system. *Phys. Chem. Chem. Phys.* **9**, 6187-6198 (2007).
38. Kitova, E.N., El-Hawiet, A., Schnier, P.D. & Klassen, J.S. Reliable Determinations of Protein–Ligand Interactions by Direct ESI-MS Measurements. Are We There Yet? *J. Am. Soc. Mass. Spectrom.* **23**, 431-441 (2012).
39. Daniel, J.M., Friess, S.D., Rajagopalan, S., Wendt, S. & Zenobi, R. Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *Int. J. Mass spectrom.* **216**, 1-27 (2002).
40. Rose, Rebecca J. et al. Quantitative Analysis of the Interaction Strength and Dynamics of Human IgG4 Half Molecules by Native Mass Spectrometry. *Structure* **19**, 1274-1282 (2011).
41. Belov, A.M. et al. Analysis of Proteins, Protein Complexes, and Organellar Proteomes Using Sheathless Capillary Zone Electrophoresis - Native Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **28**, 2614-2634 (2017).
42. VanAernum, Z.L. et al. Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. *Nat. Protoc.* **15**, 1132-1157 (2020).
43. Keifer, D.Z. & Jarrold, M.F. Single-molecule mass spectrometry. *Mass Spectrom. Rev.* **36**, 715-733 (2017).
44. Kafader, J.O. et al. Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nat. Methods* **17**, 391-394 (2020).
45. Young, G. et al. Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423-427 (2018).
46. Ortega Arroyo, J. et al. Label-Free, All-Optical Detection, Imaging, and Tracking of a Single Protein. *Nano Lett.* **14**, 2065-2070 (2014).
47. Piliarik, M. & Sandoghdar, V. Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites. *Nature Communications* **5**, 4495 (2014).
48. Cole, D., Young, G., Weigel, A., Sebesta, A. & Kukura, P. Label-Free Single-Molecule Imaging with Numerical-Aperture-Shaped Interferometric Scattering Microscopy. *ACS Photonics* **4**, 211-216 (2017).
49. Young, G. & Kukura, P. Interferometric Scattering Microscopy. *Annu. Rev. Phys. Chem.* **70**, 301-322 (2019).

50. Lai, S.-H., Tamara, S. & Heck, A.J.R. Single-particle mass analysis of intact ribosomes by mass photometry and Orbitrap-based charge detection mass spectrometry. *iScience* **24**, 103211 (2021).
51. Liebel, M., Hugall, J.T. & van Hulst, N.F. Ultrasensitive Label-Free Nanosensing and High-Speed Tracking of Single Proteins. *Nano Lett.* **17**, 1277-1281 (2017).
52. Soltermann, F. et al. Quantifying Protein–Protein Interactions by Molecular Counting with Mass Photometry. *Angew. Chem. Int. Ed.* **59**, 10774-10779 (2020).
53. Wu, D. & Piszczek, G. Measuring the affinity of protein-protein interactions on a single-molecule level by mass photometry. *Anal. Biochem.* **592**, 113575 (2020).
54. Bleeker, W.K. et al. Dual Mode of Action of a Human Anti-Epidermal Growth Factor Receptor Monoclonal Antibody for Cancer Therapy. *J. Immunol.* **173**, 4699-4707 (2004).
55. de Jong, R.N. et al. A Novel Platform for the Potentiation of Therapeutic Antibodies Based on Antigen-Dependent Formation of IgG Hexamers at the Cell Surface. *PLoS Biol.* **14**, e1002344 (2016).
56. Wu, S.-L. et al. Identification of Potential Glycan Cancer Markers with Sialic Acid Attached to Sialic Acid and Up-regulated Fucosylated Galactose Structures in Epidermal Growth Factor Receptor Secreted from A431 Cell Line*. *Mol. Cell. Proteomics* **12**, 1239-1249 (2013).
57. Cruz, A.R. et al. Staphylococcal protein A inhibits complement activation by interfering with IgG hexamer formation. *Proc. Natl. Acad. Sci.* **118**, e2016772118 (2021).
58. van Kampen, M.D. et al. Biophysical characterization and stability of IgG1 variants with different hexamerization propensities. *Work by Genmab (Utrecht, the Netherlands), submitted in August 2021* (unpublished work).
59. Gaboriaud, C. et al. Structure and activation of the C1 complex of complement: unraveling the puzzle. *Trends Immunol.* **25**, 368-373 (2004).
60. Mortensen, S.A. et al. Structure and activation of C1, the complex initiating the classical pathway of the complement cascade. *Proc. Natl. Acad. Sci.* **114**, 986-991 (2017).
61. Hughes-Jones, N.C. & Gardner, B. Reaction between the isolated globular sub-units of the complement component Clq and IgG-complexes. *Mol. Immunol.* **16**, 697-701 (1979).
62. Feinstein, A., Richardson, N. & Taussig, M.I. Immunoglobulin flexibility in complement activation. *Immunol. Today* **7**, 169-174 (1986).
63. Burton, D.R. Immunoglobulin G: Functional sites. *Mol. Immunol.* **22**, 161-206 (1985).
64. Sharp, T.H. et al. Insights into IgM-mediated complement activation based on in situ structures of IgM-C1-C4b. *Proc. Natl. Acad. Sci.* **116**, 11900-11905 (2019).

Supporting Information

Supplementary Methods

Protein samples

Anti-EGFR antibodies (2F8 mAb in IgG4Δhinge, IgG1 and IgG1-RGY format) and sEGFR were recombinantly expressed and purified by Genmab as described previously[1-4]. Purified human C1q was obtained from Complement Technology. All chemicals used were of analytical grade or higher. Protein samples were buffer exchanged to the appropriate buffer solution (150 mM ammonium acetate pH 7.5 or PBS pH 7.4) in six consecutive dilution and concentration steps at 4 °C using Amicon Ultra centrifugal filters with a 10 kDa molecular weight cutoff (Merck). Concentrations of the initial stock solutions were determined using the absorbance at 280 nm as measured by a NanoDrop 1000 spectrophotometer (NanoDrop Technologies) and the molar extinction coefficient calculated by Expasy's ProtParam. Protein complexes were assembled by mixing subcomponents at the desired molar ratios, followed by incubation at room temperature for at least 30 minutes. Anti-EGFR mAbs were incubated at 2 μM with 5 μM (MP and native MS) or 8 μM (SEC-MALS) sEGFR and/or 0.5 μM C1q. For quantitative experiments, the incubation step after preparing a dilution series proceeded for at least 4 hours to allow re-equilibration.

Native MS

Native MS experiments were performed measuring the proteins in 150 mM aqueous ammonium acetate pH 7.5. Samples were loaded into gold-coated borosilicate capillaries (prepared in-house) for direct infusion from a static nano-electrospray ionization source. Quantitative analyses of IgG1-RGY hexamers were performed on a modified LCT time-of-flight instrument (Waters), measuring each sample of a dilution series in triplicate for 1 min (30 scans) after a stable signal had been obtained. Data were processed in MassLynx V4.1 (Waters), followed by analysis using an in-house script that sums and compares ion intensities within the *m/z* ranges that we assigned to the different species. This approach is similar to a method described by Wang et al[5]. All other experiments were performed on a Q Exactive Plus UHMR Orbitrap instrument (Thermo Fisher Scientific), typically collecting at least 100 scans. The set resolution of the instrument at *m/z* 200 was 50,000 for IgG1, 6250 for sEGFR, 25,000 for IgG1 + sEGFR, 50,000 for individual measurements of IgG1-RGY and C1q, and 25,000 for complexes of IgG1-RGY, sEGFR and C1q, Data were exported from Thermo Xcalibur Qual Browser 4.2.28.14 (Thermo Fisher Scientific), and all figures were prepared using an in-house Python script in Jupyter Notebook. Masses were determined by Bayesian deconvolution using UniDec 4.4.1[6].

Native CDMS

Similar as for the normal native MS experiments, samples were introduced into a Q Exactive Plus UHMR Orbitrap mass spectrometer, though operated with a relatively low collision gas pressure (N_2). The pressure readout of the UHV sensor was controlled below $2 \cdot 10^{-10}$ torr. The instrumental resolution was set to 280,000 at 200 *m/z* (transient of 1 s) for accurate detection of both *m/z* and *z*. After multiscan acquisition, RAW files were centroided and converted into mzXML format, followed by filtering out ions that were

dephased during the transient and therefore occurred as split peaks[7]. The intensity noise threshold was set to the level corresponding to 8 elementary charges. After the filtering, the remaining centroids of single ion events are more confined in the intensity domain as well as their charge states. In this work, a calibration factor of 12.55 (normalized arbitrary intensities/charges) was used for correlating the measured intensities and charges of individual single ions. Several mzXML files could be merged into one for providing improved statistics. According to the determined charge state, a resulting formula mass = $m/z \cdot z - z$ was used to calculate the mass of every single ion. Finally, entire mass distributions were plotted and fitted with a normal distribution in the histogram (mass and counts) with an appropriate bin size (10 kDa).

SEC-MALS

SEC-MALS data were collected using a Waters HPLC with an in-line UV detector (Waters 2487 Dual Absorbance), a MALS detector (MiniDAWN, Wyatt Technology) and an RI detector (Optilab, Wyatt Technology). Analysis was performed on an SRT SEC-500 column (Sepax Technologies) using 100 mM sodium phosphate, 100 mM sodium sulfate, pH 6.8 as mobile phase at 0.35 mL/min flow rate. By simultaneously measuring the light scattering and the concentration of the molecules as they elute from the column, the molar mass of the particle could be determined. As concentration determination by UV absorbance requires prior knowledge of the extinction coefficient, this brings SEC-MALS to a problem of circularity when the aim is to determine the stoichiometry of a complex using only a UV detector as concentration source. This problem is overcome by using a refractive index (RI) detector, which can assess concentrations directly. MALS data were processed by ASTRA software (Wyatt version 8.00.25) based on MALS-RI for antibody mass determinations (using the RI detector as concentration source with a *dn/dc* of 0.185) and with the Protein Conjugate Analysis method (MALS-UV-RI) to get information on the individual contributions of a) glycan and protein for the glycosylated proteins and b) different protein subunits for the complexes. For data processing of the individual glycosylated proteins, *dn/dc* values of 0.185 and 0.140 were used as input for "protein" and "modifier" (glycan) portions respectively. The complexes were processed in a multi-step approach essentially according to Hastie *et al*.[8]. First, the antibody-antigen complex was analyzed using the Protein Conjugate Analysis method with the antibody assigned as "protein" and the antigen (sEGFR) as "modifier". For the antibody, a *dn/dc* of 0.185 and the UV extinction coefficient as calculated by ASTRA from MALS-RI analysis of the individual antibody were entered as "protein" parameters. For the antigen, these values were derived from MALS-UV-RI analysis of the individual compound, as calculated by ASTRA, and entered as "modifier" parameters. The three-component C1q-antibody-sEGFR complex was analyzed taking the calculated UV extinction coefficient and *dn/dc* values from the previous steps as input values. Results are reported as molar mass at the UV peak maximum (M_p).

Mass photometry

MP experiments were performed on a Refeyn One^{MP} mass photometer (Refeyn), measuring all samples in PBS. Borosilicate microscope coverslips (24 x 50 mm 1.5H, Marienfeld) were cleaned in four sequential sonication rounds of 5 min using isopropanol and MilliQ water (2x). Silicone cell culture gaskets (50 wells, 3 mm diameter x 1 mm depth,

Grace Bio-Labs) were cut into sets of four wells, which were placed onto clean coverslips. Before data acquisition, coverslip wells were loaded with approximately 12 µL PBS, followed by focusing the instrument on the glass-liquid interface. All samples were measured in triplicate. Typically, samples were added by pipetting 3 µL of a diluted solution directly into the PBS-loaded well to a final concentration of 1-30 nM, followed by recording for 150 s (12,000 frames). When measuring protein complexes, higher concentration solutions in the µM range were jump diluted to nM range concentrations, starting the recording within 5-30 s from the initial dilution step. Recordings were processed in DiscoverMP (Refeyn) and calibrated using an in-house calibration mix consisting of proteins of which accurate masses had previously been determined by MS (73 kDa IgG4Δhinge-L368A, 149 kDa IgG1-Campath, 483 kDa apoferritin, and 800 kDa GroEL). Data analysis and plotting were performed in Jupyter Notebook using an in-house Python library, combining multiple measurements of the same sample into a single dataset. For quantitative MP experiments, we jump diluted samples from a dilution series in triplicate and recorded for 75 s (6000 frames). Dissociation of protein complexes upon jump dilution was modelled to infer their abundance in the original solution (see below).

Modelling IgG4Δhinge dissociation to estimate the fractional occupancy before jump dilution using MP

Determining K_d values works best when the sample is analyzed at concentrations around the K_d value of the analyte, the point at which 50% is bound. MP experiments are however mostly performed at nM range concentrations, limiting the ability to assess µM-mM range affinities. This problem can be partially overcome by jump dilution just before measurement, as the distribution of protein species then roughly corresponds to the equilibrium before dilution. However, this method only works if the k_{off} is relatively low. In practice, when k_{off} is not low, the complex can dissociate during the measurement, which needs to be corrected for. Our solution to this problem is to infer the k_{off} of the interaction from the observed decay of protein complexes over time, which can then be used to estimate the complex abundance at the instant of dilution (t = 0).

For IgG4Δhinge, the concentration of $(HL)_2$ over time can be described using the following equation:

$$[(HL)_2] = f(t) = \left([(HL)_2]_0 - [(HL)_2]_{eq}\right) \cdot e^{-\left(k_{on}\cdot[HL]+k_{off}\right)\cdot t} + [(HL)_2]_{eq} \quad \text{(S1)}$$

In this formula, $[(HL)_2]_0$ is the concentration of $(HL)_2$ at the instant of jump dilution, while $[(HL)_2]_{eq}$ is the equilibrium concentration under measurement conditions. However, when measuring at concentrations far below the K_d, we can assume that $[(HL)_2]_{eq}$ and k_{on} * [HL] are close to zero:

$$[(HL)_2] = f(t) = [(HL)_2]_0 \cdot e^{-k_{off}\cdot t} \quad \text{(S2)}$$

The fractional occupancy corresponding to $[(HL)_2]_0$ is then identical to the fractional occupancy at the concentration before dilution (t=-1). Therefore:

(S3) $$[(HL)_2]_0 = \frac{FO_{-1} \cdot [(HL)_{total}]_0}{2}$$

With:

(S4) $$FO = \frac{2 \cdot [(HL)_2]}{[(HL)_{total}]}$$

Only $[(HL)_{total}]_{-1}$ is known. However, we can write $[(HL)_2]$ as a function of $[(HL)_{total}]$ and the K_d:

(S5) $$K_d = \frac{[A][B]}{[AB]} = \frac{[HL][HL]}{[(HL)_2]}$$

(S6) $$K_d = \frac{([(HL)_{total}] - 2[(HL)_2])^2}{[(HL)_2]}$$

(S7) $$[(HL)_2]K_d = ([(HL)_{total}] - 2[(HL)_2])^2$$

(S8) $$[(HL)_2]K_d = [(HL)_{total}]^2 - 4[(HL)_2][(HL)_{total}] + 4[(HL)_2]^2$$

(S9) $$4[(HL)_2]^2 - (4[(HL)_{total}] + K_d)[(HL)_2] + [(HL)_{total}]^2 = 0$$

Which has the form:

(S10) $$ax^2 + bx + c = 0$$

According to the quadratic equation:

(S11) $$[(HL)_2] = \frac{(4[(HL)_{total}] + K_d) - \sqrt{(4[(HL)_{total}] + K_d)^2 - 16[(HL)_{total}]}}{8}$$

Therefore $[(HL)_2]_0$ is:

(S12) $$[(HL)_2]_0 = \frac{FO \cdot [(HL)_{total}]_0}{2} = \frac{\frac{2 \cdot [(HL)_2]_{-1}}{[(HL)_{total}]_{-1}} \cdot [(HL)_{total}]_0}{2}$$

(S13) $$[(HL)_2]_0 = \frac{\frac{2 \cdot \left(4[(HL)_{total}]_{-1} + K_d\right) - \sqrt{\left(4[(HL)_{total}]_{-1} + K_d\right)^2 - 16[(HL)_{total}]_{-1}}}{8}}{[(HL)_{total}]_{-1}} \cdot [(HL)_{total}]_0}{2}$$

(S14) $$[(HL)_2]_0 = \frac{(4[(HL)_{total}]_{-1}+K_d)-\sqrt{(4[(HL)_{total}]_{-1}+K_d)^2-16[(HL)_{total}]_{-1}}}{8[(HL)_{total}]_{-1}} \cdot [(HL)_{total}]_0$$

And the concentration of $(HL)_2$ over time is:

(S15) $$[(HL)_2] = \frac{(4[(HL)_{total}]_{-1}+K_d)-\sqrt{(4[(HL)_{total}]_{-1}+K_d)^2-16[(HL)_{total}]_{-1}}}{8[(HL)_{total}]_{-1}} \cdot [(HL)_{total}]_0 \cdot e^{-k_{off}\cdot t}$$

Alternatively, we can describe FO over time:

(S16) $$FO = \frac{(4[(HL)_{total}]_{-1}+K_d)-\sqrt{(4[(HL)_{total}]_{-1}+K_d)^2-16[(HL)_{total}]_{-1}}}{4[(HL)_{total}]_{-1}} \cdot e^{-k_{off}\cdot t}$$

The unknown k_{off} can then be estimated by fitting the above formula to the measurement of FO over time. This also gives a first indication of the K_d. However, fitting can best be done on a linear scale, which is why we used ln(FO).

(S17) $$\ln(FO) = \ln\left(\frac{(4[(HL)_{total}]_{-1}+K_d)-\sqrt{(4[(HL)_{total}]_{-1}+K_d)^2-16[(HL)_{total}]_{-1}}}{4[(HL)_{total}]_{-1}}\right) + \ln\left(e^{-k_{off}\cdot t}\right)$$

(S18) $$\ln(FO) = \ln\left(\frac{(4[(HL)_{total}]_{-1}+K_d)-\sqrt{(4[(HL)_{total}]_{-1}+K_d)^2-16[(HL)_{total}]_{-1}}}{4[(HL)_{total}]_{-1}}\right) - k_{off}\cdot t$$

In order to determine the k_{off}, high-concentration samples of each IgG4Δhinge variant were jump diluted and measured for 10 minutes in triplicate, pipetting the solution up and down every two minutes to keep a sufficient detection rate (**Figure S1**). For assemblies with high k_{off} values, we only considered the first 100 s of the measurement, as the decay rate of FO already started to level off as the sample approached equilibrium. Equation S18 was then fitted to the data using linear least squares regression, readily revealing k_{off} values spanning three orders of magnitude.

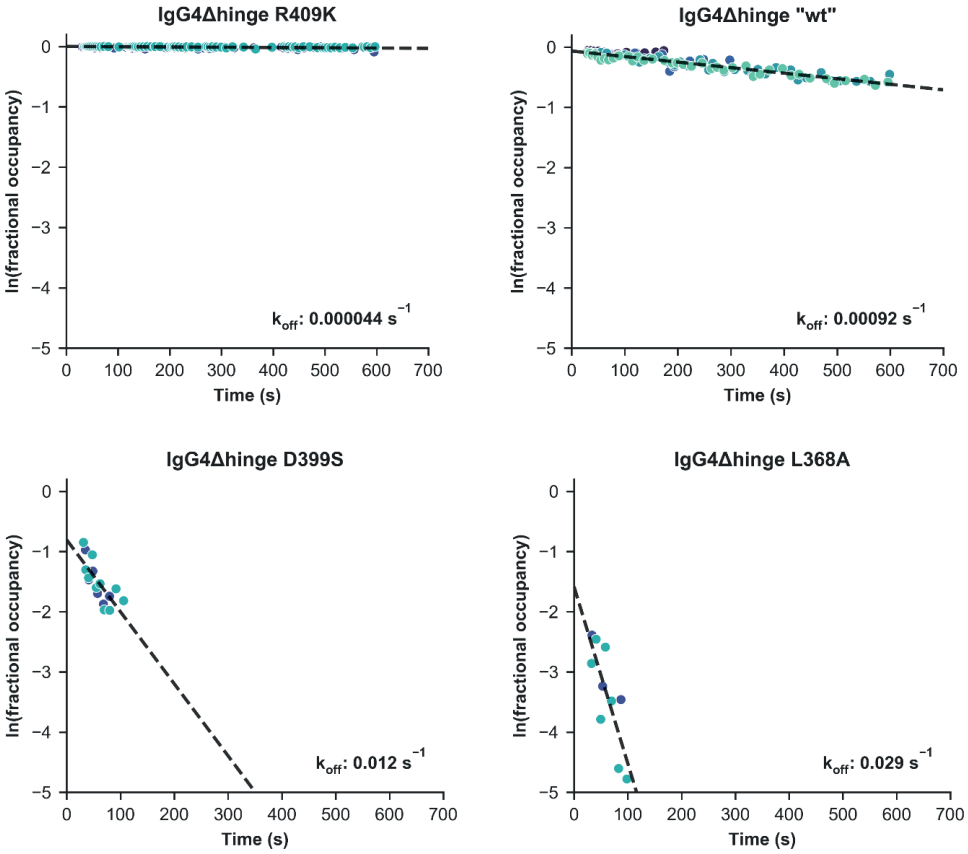


Figure S1 | Using MP to determine the k_{off} of IgG4Δhinge variants by following the fractional occupancy of the dimer over time. High-concentration samples were jump diluted and measured for 10 minutes, pipetting the solution up and down every two minutes to maintain a sufficient detection rate. Events were assigned to (HL) or $(HL)_2$ based on a mass window corresponding to an equal portion of the normal distribution (*e.g.*, from μ - 2σ to μ + 2σ). Assigned events were grouped into bins of 100, followed by calculating the fractional occupancy within the bin (colored for different replicates). Equation S18 was fitted to the data using linear least squares regression to estimate the k_{off} of the interaction (black line).

Measuring affinities by dissociation-corrected jump dilution MP

The K_d of IgG4Δhinge variants was determined by measuring a dilution series in triplicate using a standard recording length of 75 s (6000 frames). To correct for jump dilution-induced dissociation, we fitted equation S18 with the previously determined k_{off} to each measurement to in infer the fractional occupancy at t=0. Fits were checked manually, and poor fits were removed upon inspection of the original data, for example when too few binding events were available. Furthermore, because K_d determination gets increasingly inaccurate at the extremes of fractional occupancy (nearly 0% or 100% dimer), where possible, we only considered data points where 5% < FO < 95%.

Using the estimated fractional occupancy, [HL] and [$(HL)_2$] can be calculated. The ratio between these concentrations gives the K_d:

(S19) $$K_d = \frac{[HL][HL]}{[(HL)_2]}$$

(S20) $$[HL]^2 = [(HL)_2] \cdot K_d$$

In a plot of $[HL]^2$ versus [$(HL)_2$], the slope thus gives the K_d. However, the concentration points in the dilution series were not distributed on a linear scale, but rather a logarithmic scale. To avoid overweighting of the highest concentration points, we therefore proceeded with a log transform of the dataset:

(S21) $$\log([HL]^2) = \log([(HL)_2] \cdot K_d) = \log([(HL)_2]) + \log(K_d)$$

We used least squares fitting of equation S21 to the determine the overall K_d, which in a plot of $\log([HL]^2)$ versus $\log([(HL)_2])$ is defined by the intersect with the y-axis (**Figure S2**).

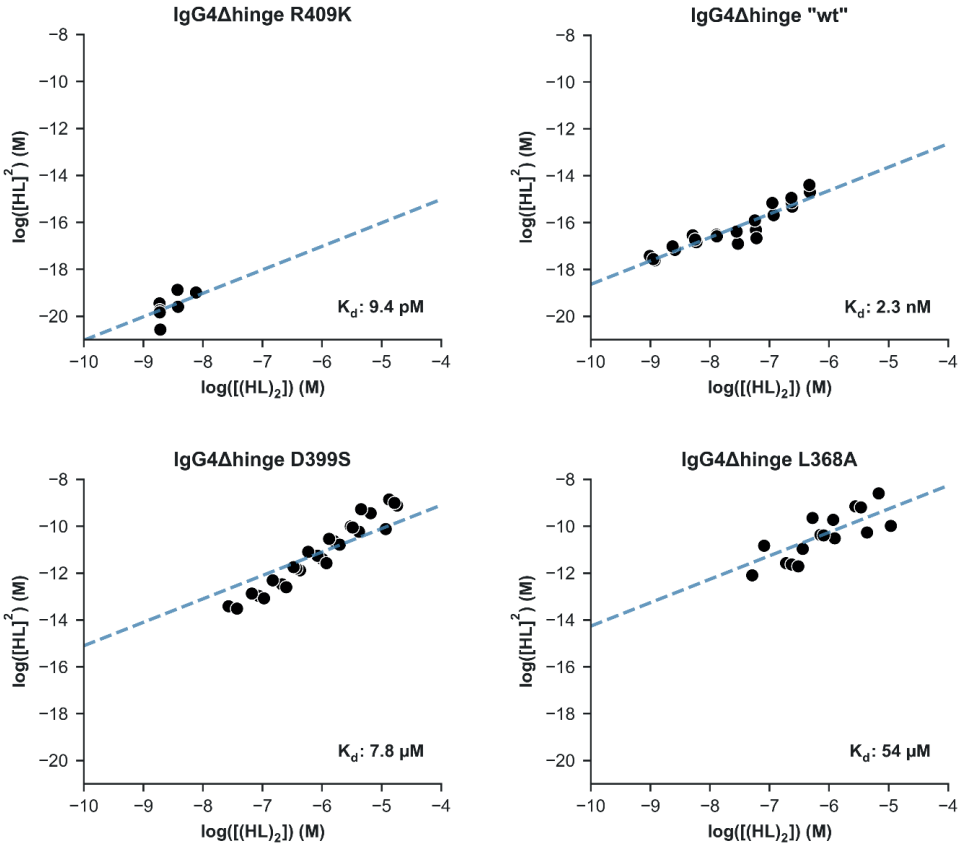


Figure S2 | Assessing the K_d of IgG4Δhinge variants from a dilution series by dissociation-corrected MP. A dilution series of IgG4Δhinge variants was measured by jump dilution MP in triplicate, followed by inference of the fractional occupancy at t=0 using the previously determined k_{off}. Fits were checked manually, outliers were removed, and when possible, we only considered measurements where 5% < FO < 95%. The apparent K_d of each measurement was then determined by linear least squares fitting of equation S21 (blue line), revealing the overall K_d value.

Supplementary Figures

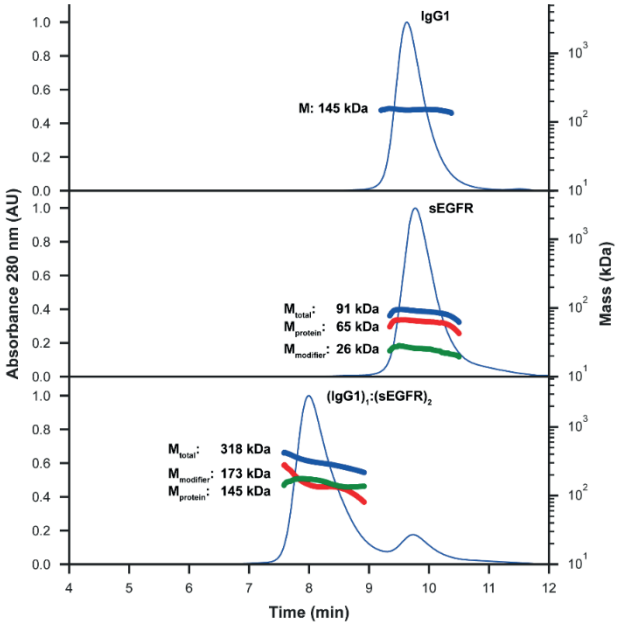


Figure S3 | Characterizing highly heterogeneous antibody-antigen complexes involving IgG1 and sEGFR by SEC-MALS. Using the RI detector for direct concentration measurement, SEC-MALS quite accurately determined the mass of IgG1 anti-sEGFR. For more complex samples, SEC-MALS-UV-RI "protein-conjugate" analysis enabled disentanglement of the mass contributions of a protein backbone (based on the UV absorption coefficient) and a modifier (glycans or a binding partner) from the total mass. This approach enabled total mass determination for sEGFR, while simultaneously analyzing the contribution of the protein backbone (protein) and glycans (modifier). When 2 µM of IgG1 was incubated with 8 µM of sEGFR, similarly, the mass of the full $(IgG1)_1$:$(sEGFR)_2$ complex could be determined, disentangling the contributions from the IgG1 (protein) and two bound sEGFR molecules (modifier).

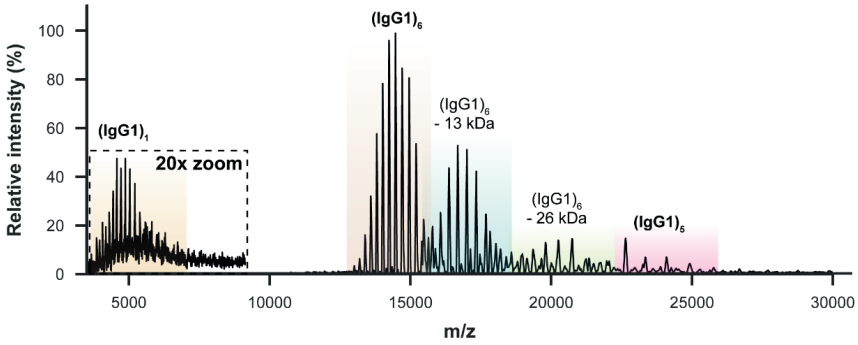


Figure S4 | Gas phase activation of IgG1-RGY hexamers leads to ejection of a highly charged IgG1 monomer. The charge state envelope of $(IgG1)_6$ was isolated by the quadrupole of the instrument, followed by collisional activation using Xenon gas in the HCD cell. This led to the ejection of a highly charged monomer, concomitant with a lowly charged pentamer. We additionally observed the ejection of 13 kDa polypeptides from the hexameric complex, likely resulting from a specific backbone cleavage in the Fab portion of the antibodies.

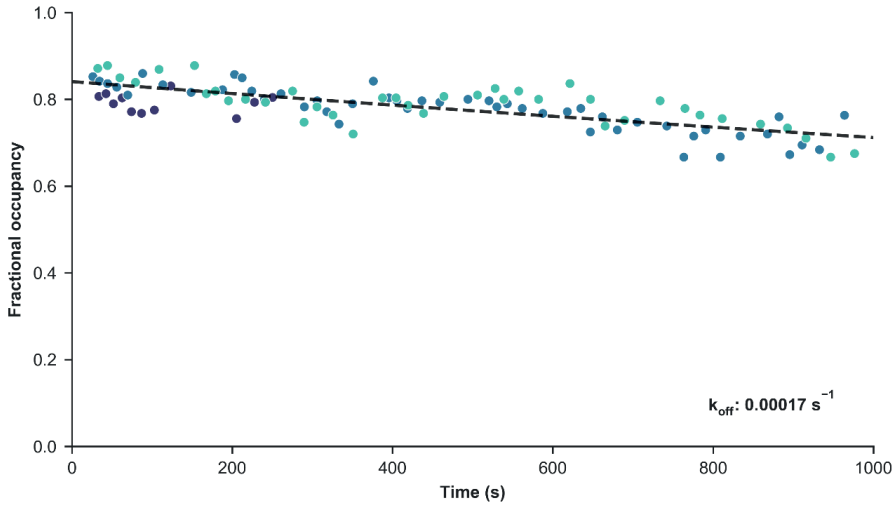


Figure S5 | IgG1-RGY hexamers dissociate slowly upon jump dilution MP. IgG1-RGY samples of 8 µM were jump diluted to a measurement concentration of 10 nM, followed by a MP recording of 10 minutes, pipetting the solution up and down every two minutes to maintain a sufficient detection rate. Events were assigned to $(IgG)_1$ or $(IgG)_6$ based on a mass window corresponding to an equal portion of the normal distribution (µ - 2σ to µ + 2σ). Assigned events were grouped into bins of 100, followed by calculating the fractional occupancy within the bin (colored for three different replicates). An exponential decay function was fitted to linearized (logarithmically transformed) data using linear least squares regression to estimate the overall k_{off} of the interaction (black line).

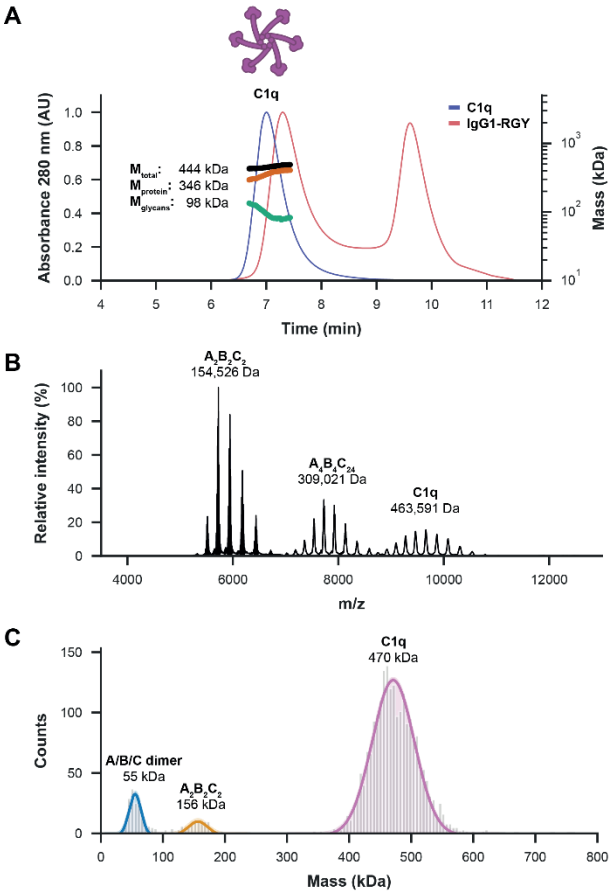


Figure S6 | The complement recognition complex C1q is reliably mass analyzed by MP while analyses by native MS and SEC are hampered by technique-related artifacts. (A) The SEC chromatogram of C1q reveals one elution peak at a relatively short retention time, suggesting a > 1 MDa particle, while MALS reveals a mass of 444 kDa, closely matching C1q. SEC-MALS-UV-RI analysis based on the calculated UV extinction coefficient revealed contributions of 346 kDa from the protein backbone and 98 kDa from the glycans. For comparison, the chromatogram of IgG-RGY is overlayed. **(B)** While native MS measures an accurate mass of 464 kDa for the intact C1q complex, it reveals the presence of 2-, 4-, and 6-armed versions of C1q. These smaller complexes likely result from dissociation induced in the source region. **(C)** Characterization of C1q by MP reveals a main mass distribution of particles corresponding to the full 6-armed C1q complex with a smaller contribution being made by a 2-armed complex. Masses closely match those observed by native MS.

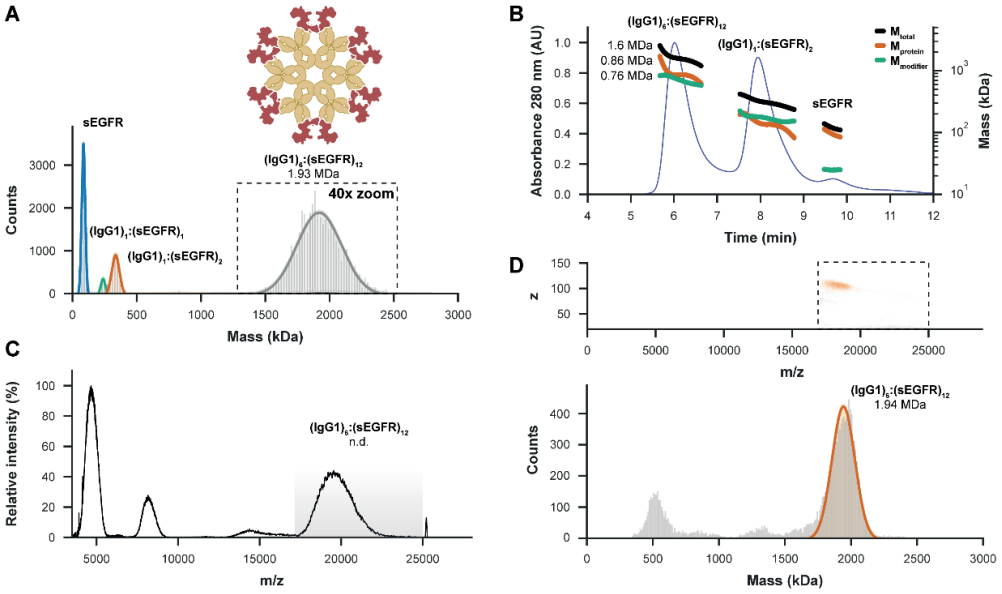


Figure S7 | MP and CDMS successfully determine the mass and full occupancy of highly heterogeneous $(IgG1)_6$:$(sEGFR)_{12}$ antibody-antigen complexes. A. MP measurements of anti-EGFR IgG1-RGY hexamers incubated with sEGFR reveal the presence of particles with an average mass of 1.93 MDa, likely corresponding to $(IgG1)_6$:$(sEGFR)_{12}$. **B.** SEC-MALS-UV-RI reveals the presence of larger immune complexes with a total estimated mass of around 1.6 MDa (~ 0.86 MDa for $(IgG1)_6$ with a ~ 0.76 MDa modifier, *i.e.*, sEGFR molecules). The mass accuracy of SEC-MALS is in this case insufficient for determining the exact stoichiometry of $(IgG1)_6$:(sEGFR) complexes. **C.** Native mass spectra of the same sample show that although large ion species are detected at about *m/z* 20,000, they cannot be charge-resolved, preventing mass determination. **D.** Two-dimensional charge-detection MS spectra reveal the *m/z* and *z* of single ions independently, circumventing the need for charge-state resolution. For the main particle population an average mass of 1.94 MDa was determined, corresponding to the fully occupied $(IgG1)_6$:$(sEGFR)_{12}$ complex (bottom).

Supplementary Tables

Table S1 | Comparison of mass measurements of proteins and protein complexes between native MS, CDMS, MP and SEC-MALS. For native MS, when individual proteoglycoforms could be resolved, the intensity-weighted average mass is indicated. For MP, masses indicate the mean of a normal distribution fitted to a range of single-particle mass measurements. For CDMS, masses indicate the maximum of a kernel density estimation. Masses measured by MP and SEC-MALS are compared to highly accurate masses determined by native MS. For complexes involving sEGFR, the expected mass was calculated from the sum of components using a previously reported tandem MS-derived mass for sEGFR[9]. These expected values are given in italics and in parentheses.

Species	$MW_{Native\ MS}$ (kDa)	MW_{CDMS} (kDa)	ΔMW_{CDMS} (%)	MW_{MP} (kDa)	ΔMW_{MP} (%)	$MW_{SEC\text{-}MALS}$ (kDa)	$\Delta MW_{SEC\text{-}MALS}$ (%)
IgG4Δhinge HL	73.196			74	1.1%	-	-
IgG4Δhinge $(HL)_2$	146.358			146	-0.2%	-	-
sEGFR	*(87.5)*	87.7	0.2%	86	-1.7%	91	4.0%
IgG1 anti-sEGFR	149.243	146	-2.2%	148	-0.8%	145	-2.8%
$(IgG1)_1$:$(sEGFR)_1$	*(236.7)*	241	1.8%	231	-2.4%	-	-
$(IgG1)_1$:$(sEGFR)_2$	*(324.2)*	327	0.9%	320	-1.3%	318	-1.9%
IgG1-RGY	149.607			148	-1.1%	160	6.9%
$(IgG1\text{-}RGY)_6$	897.645			965	7.5%	818	-8.9%
$(sEGFR)_{12}$:$(IgG1\text{-}RGY)_6$	*(1,947.6)*	1,942	-0.3%	1,929	-1.0%	1621	-16.8%
C1q	463.561			470	1.4%	444	-4.2%
$(IgG1\text{-}RGY)_6$:$(A_4B_4C_4)$	1,206.181			-	-	-	-
$(IgG1\text{-}RGY)_6$:C1q	1,361.042			1,425	4.7%	1252	-8.0%
$(sEGFR)_{12}$:$(IgG1\text{-}RGY)_6$:C1q	*(2,411.2)*	2,422	0.4%	2,354	-2.4%	1944	-19.4%

Table S2 | MP-derived kinetic rates and equilibrium constants for IgG4Δhinge variants. Dissociation rates for IgG4Δhinge variants were determined by following the fractional occupancy over time for jump-diluted high-concentration samples. A dilution series of each mutant was recorded and corrected for dissociation upon jump dilution using the determined k_{off} values. This enabled us to estimate the fractional occupancy in the original samples and to determine the K_d of the interaction. The k_{on} was subsequently calculated by dividing the k_{off} by the K_d. In the last column, the values determined by native MS as previously reported by Rose *et al.*[1] are given.

IgG4Δhinge variant	k_{off} (s^{-1})	k_{on} ($M^{-1}\ s^{-1}$)	K_d (M)	$K_{d,\ native\ MS}$ (M)[1]
R409K	$4.4 \cdot 10^{-5}$	$4.7 \cdot 10^{6}$	$9.4 \cdot 10^{-12}$	$4.0 \cdot 10^{-10}$
Wild-type	$9.2 \cdot 10^{-4}$	$3.1 \cdot 10^{5}$	$2.9 \cdot 10^{-9}$	$5.0 \cdot 10^{-8}$
D399S	$1.8 \cdot 10^{-2}$	$2.3 \cdot 10^{3}$	$7.8 \cdot 10^{-6}$	$4.4 \cdot 10^{-6}$
L368A	$3.0 \cdot 10^{-2}$	$5.5 \cdot 10^{2}$	$5.4 \cdot 10^{-5}$	$7.6 \cdot 10^{-6}$

Table S3 | Advantages and disadvantages of MP, native MS and SEC-MALS in the analysis of antibodies and heavily glycosylated macromolecular immune complexes.

	Advantages	Disadvantages
MP	• Fastest • Most native conditions • Low sample consumption • Requires little training • Accurate quantification at nM range protein concentrations	• Limited mass resolution and accuracy • Not useful for interactions causing a small mass shift • Mass artifacts for some proteins • Limited to strongly bound and stable interactions at nM concentrations • Large excesses of certain components complicate the concomitant detection of particles at lower abundance • Available methodologies less robust in dilution and sample handling
Native MS	• Supreme mass resolution and accuracy • Accurate quantification at µM range protein concentrations • Workflows available for automated sample handling and high-throughput data collection	• Requires training • Requires buffer exchange • Hampered by extensive sample microheterogeneity • ESI process may induce dissociation for some protein complexes
Charge detection native MS	• Low sample consumption • Supreme mass resolution and accuracy also for highly heterogenous samples	• Requires extensive training • Not well established yet
SEC-MALS	• Non-destructive • Reliable quantification of protein abundance • Well-established for one-component systems and simple protein complexes • Highly robust workflows available for automated sample handling and high-throughput data collection	• Low resolution and mass accuracy • Not useful for interactions causing a small mass shift • Limited to slowly dissociating interactions • Highest sample consumption

Supplementary References

1. Rose, Rebecca J. et al. Quantitative Analysis of the Interaction Strength and Dynamics of Human IgG4 Half Molecules by Native Mass Spectrometry. *Structure* **19**, 1274-1282 (2011).
2. Bleeker, W.K. et al. Dual Mode of Action of a Human Anti-Epidermal Growth Factor Receptor Monoclonal Antibody for Cancer Therapy. *J. Immunol.* **173**, 4699-4707 (2004).
3. de Jong, R.N. et al. A Novel Platform for the Potentiation of Therapeutic Antibodies Based on Antigen-Dependent Formation of IgG Hexamers at the Cell Surface. *PLoS Biol.* **14**, e1002344 (2016).
4. Wang, G. et al. Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Mol. Cell* **63**, 135-145 (2016).
5. Wang, G., Johnson, A.J. & Kaltashov, I.A. Evaluation of Electrospray Ionization Mass Spectrometry as a Tool for Characterization of Small Soluble Protein Aggregates. *Anal. Chem.* **84**, 1718-1724 (2012).
6. Marty, M.T. et al. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370-4376 (2015).
7. Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat Methods* **17**, 395-398 (2020).
8. Hastie, K., Rayaprolu, V. & Saphire, E.O. in Mass Spectrometry of Glycoproteins: Methods and Protocols. (ed. A. Delobel) 343-359 (Springer US, New York, NY; 2021).
9. Wang, G., de Jong, R.N., van den Bremer, E.T.J., Parren, P.W.H.I. & Heck, A.J.R. Enhancing Accuracy in Molecular Weight Determination of Highly Heterogeneously Glycosylated Proteins by Native Tandem Mass Spectrometry. *Anal. Chem.* **89**, 4793-4797 (2017).

CHAPTER 4

STAPHYLOCOCCAL PROTEIN A INHIBITS COMPLEMENT ACTIVATION BY INTERFERING WITH IGG HEXAMER FORMATION

Ana Rita Cruz[1,*], Maurits A. den Boer[2,3,*], Jürgen Strasser[3], Seline A. Zwarthoff[1], Frank J. Beurskens[5], Carla J.C. de Haas[1], Piet C. Aerts[1], Guanbo Wang[2,3,6], Rob N. de Jong[5], Fabio Bagnoli[7], Jos A.G. van Strijp[1], Kok P.M. van Kessel[1], Janine Schuurman[5], Johannes Preiner[4], Albert J.R. Heck[2,3,#], and Suzan H.M. Rooijakkers[1,#]

[1] Medical Microbiology, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

[2] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

[4] University of Applied Sciences Upper Austria, 4020 Linz, Austria

[5] Genmab, 3508 AD Utrecht, The Netherlands

[6] School of Chemistry and Materials Science, Nanjing Normal University, Nanjing 210023, China

[7] GSK Siena, Italy

[*] These authors contributed equally

[#] These authors shared supervision of this study

Published as:

Staphylococcal protein A inhibits complement activation by interfering with IgG hexamer formation
Ana Rita Cruz, Maurits A. den Boer, Jürgen Strasser, Seline A. Zwarthoff, Frank J. Beurskens, Carla J.C. de Haas, Piet C. Aerts, Guanbo Wang, Rob N. de Jong, Fabio Bagnoli, Jos A.G. van Strijp, Kok P.M. van Kessel, Janine Schuurman, Johannes Preiner, Albert J.R. Heck, and Suzan H.M. Rooijakkers
Proceedings of the National Academy of Sciences 2021 *118* (7) e2016772118
DOI: 10.1073/pnas.2016772118

Significance

Antibodies are crucial for the immune response against bacteria. To drive bacterial killing, antibodies should bind to the bacterial cell and induce the complement reaction. This requires target-bound IgGs to form hexameric IgG platforms that are kept together by non-covalent Fc-Fc interactions. Interestingly, pathogenic bacteria produce IgG-binding molecules that bind exactly to the Fc region needed for hexamerization. Here we demonstrate that staphylococcal protein A (SpA) from *Staphylococcus aureus* specifically blocks the formation of IgG hexamers and downstream activation of complement. Furthermore, we show that IgG3 antibodies (which are not recognized by SpA) have a superior capacity to activate complement and induce the killing of *S. aureus* by human phagocytes. These insights provide a crucial rationale for optimizing antibody therapies against *S. aureus*.

Abstract

Immunoglobulin (Ig) G molecules are essential players in the human immune response against bacterial infections. An important effector of IgG-dependent immunity is the induction of complement activation, a reaction that triggers a variety of responses that help kill bacteria. Antibody-dependent complement activation is promoted by the organization of target-bound IgGs into hexamers that are held together via noncovalent Fc-Fc interactions. Here we show that staphylococcal protein A (SpA), an important virulence factor and vaccine candidate of *Staphylococcus aureus*, effectively blocks IgG hexamerization and subsequent complement activation. Using native mass spectrometry and high-speed atomic force microscopy, we demonstrate that SpA blocks IgG hexamerization through competitive binding to the Fc-Fc interaction interface on IgG monomers. In concordance, we show that SpA interferes with the formation of $(IgG)_6$:C1q complexes and prevents downstream complement activation on the surface of *S. aureus.* Finally, we demonstrate that IgG3 antibodies against *S. aureus* can potently induce complement activation and opsonophagocytic killing even in the presence of SpA. Together, our findings identify SpA as an immune evasion protein that specifically blocks IgG hexamerization.

Introduction

Antibodies play a key role in the human immune response against bacterial infections. While antibodies can bind and neutralize bacterial virulence factors, they can also signal to components of the innate immune system and induce bacterial killing. To do so, antibodies bind bacterial cells via their variable (Fab) region and subsequently trigger Fc-mediated effector functions[1]. The complement system, a large network of plasma proteins, forms an important effector of antibody-dependent immune protection against invading bacteria. An activated complement cascade results in the efficient decoration of bacteria with C3-derived molecules that are essential to trigger highly effective phagocytic uptake via complement receptors on phagocytes. Furthermore, complement generates chemoattractants and induces direct killing of gram-negative bacteria. Because effective complement activation is an important effector mechanism of therapeutic antibodies in cancer[2], the ability of complement to kill bacteria could also be exploited for antibacterial therapies against (antibiotic-resistant) pathogens[3-5].

The antibody-driven, "classical" complement pathway is initiated when circulating C1 complexes are recruited to antibody-labeled target surfaces[6]. The most abundant antibody isotype in serum is immunoglobulin (Ig) G, which is subdivided into subclasses IgG1, IgG2, IgG3, and IgG4 in order of decreasing abundance. IgG antibodies can bind surface antigens via their Fab regions and subsequently recruit C1 via their Fc region (**Figure S1A**). The C1 complex consists of three large units: C1q, C1r, and C1s. C1q comprises the antibody recognition unit of the C1 complex and is composed of six globular heads connected by collagen-like stalks. On binding of C1q, its associated proteases C1r and C1s are activated to cleave other complement proteins that together form enzymes on the surface that catalyze the covalent deposition of C3b molecules onto the bacterial surface (**Figure S1A**). C3b molecules are recognized by complement receptors on phagocytes (neutrophils, macrophages), which engulf and digest bacteria intracellularly. The deposition of C3b also results in the amplification of the complement cascade and activation of downstream complement effector functions.

In recent years, it has become clear that efficient binding of C1 to target-bound IgG molecules requires IgGs to form ordered hexameric ring structures[7,8]. Cryo-electron tomography and atomic force microscopy studies revealed that the six globular heads of C1q can simultaneously bind to each of the six IgG molecules that form a hexameric binding platform[7] (**Figure S1B**). The formation of these hexamers is induced by antibody binding to surface-bound antigens and driven by noncovalent interactions between the Fc regions of neighboring IgG molecules[9] (**Figure 1A**).

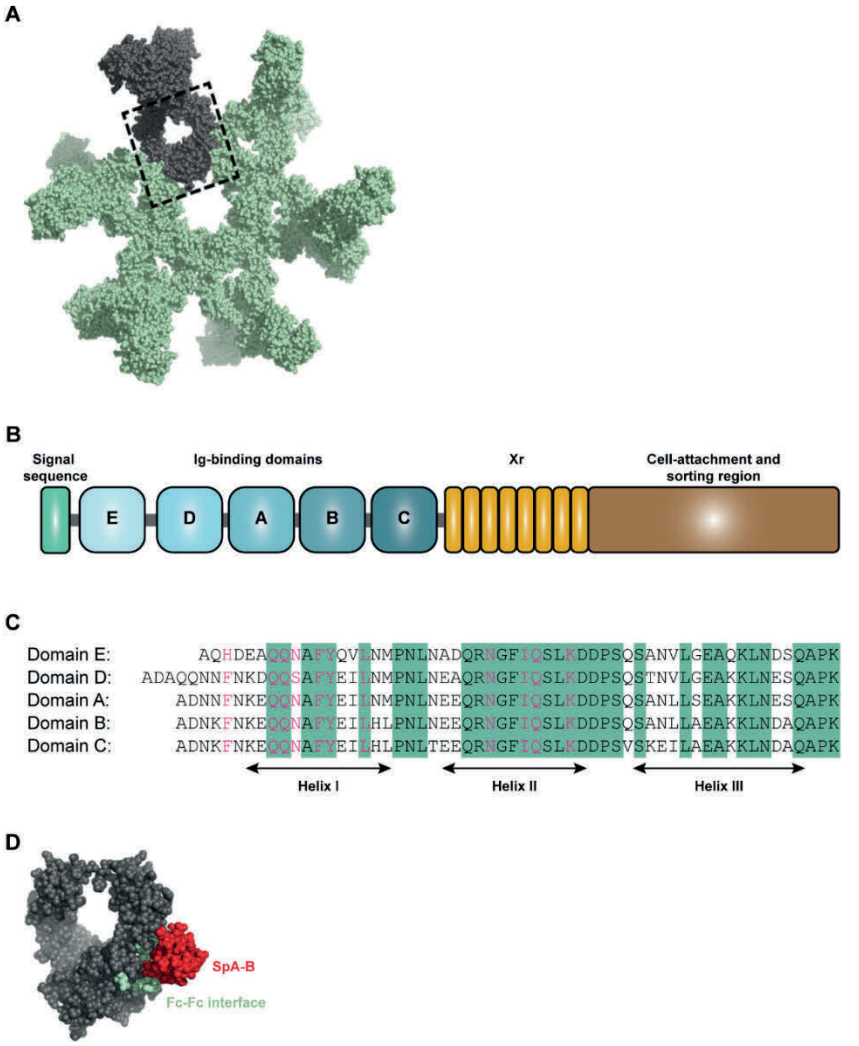


Figure 1 | The Ig-binding domains of staphylococcal protein A (SpA) bind to residues of the IgG-Fc region that are involved in IgG hexamerization. (A) IgG hexamer crystal packing of IgG1-b12 (pdb 1HZH). A single IgG is depicted in grey and IgG-Fc domain is enclosed in the dashed box. **(B)** Schematic representation of SpA organization. SpA consists of a signal sequence, five immunoglobulin (Ig)-binding domains (E, D, A, B, and C), an Xr region (octapeptide repeats variable in number), and a cell-wall attachment and sorting region that includes a constant Xc region, the LPETG motif, a hydrophobic anchor and positively charged residues. **(C)** Sequence alignment of the five highly homologous Ig-binding domains of SpA. The amino acid residues conserved in all five domains are highlighted in green. The residues involved in the interaction with the Fc region of IgG are colored in pink. **(D)** Space-filling presentation depicting the Fc domain of IgG1-b12 and its interaction with the B-domain of SpA (SpA-B: pdb 1FC2; complementary Fc docking domain for SpA-B is hidden). The residues involved in Fc-Fc interactions required to form the IgG hexameric ring are depicted in green and the crystal structure of SpA-B is shown in red.

Interestingly, some bacteria produce IgG-binding molecules that recognize the Fc domain of IgGs[10]. The best known of these is staphylococcal protein A (SpA), a 42-kDa protein that has a high affinity for the Fc region of IgG and thus is commonly used as a tool in affinity chromatography to purify monoclonal antibodies. SpA is produced by *Staphylococcus aureus*, an important human pathogen that is the main cause of serious hospital-acquired infections, such as bacteremia, sepsis, and endocarditis[11]. Due to the dramatic increase in antibiotic resistance and the lack of proper vaccines, physicians are frequently left with no useful or suboptimal alternatives when treating these infections.

SpA is considered an important virulence factor[12,13] and vaccine candidate[14,15]. The protein is abundantly present on the bacterial cell wall[16-18] but is also released in the extracellular environment[19,20]. SpA is composed of a signal sequence, five sequential Ig-binding domains (denoted E, D, A, B, and C), an Xr region and a cell wall attachment and sorting region[21] (**Figure 1B**). Each of the five repeating Ig-binding domains adopts a three-helical structure that can bind to the Fc region of IgG via helices I and II[22] and the Fab region of the VH3 type family of antibodies via helices II and III[23-25]. The binding of SpA to Ig-Fc regions is considered to protect *S. aureus* from phagocytic killing[14] while cross-linking of Ig-Fab regions triggers the proliferation and apoptotic collapse of B cells[26]. The five Ig-binding domains are highly homologous, sharing 74 to 91% of their amino acid sequence relative to the A domain[27] (**Figure 1C**). It has been demonstrated that the binding interface of the B domain of SpA (SpA-B) and IgG1-Fc involves 11 amino acid residues of SpA-B and 9 residues of IgG1-Fc[22]. Interestingly, SpA binds to all IgG subclasses except IgG3, due to a substitution in one of the nine Fc-contact residues in IgG3 (His^{435} in IgG1 becomes Arg^{435} in IgG3)[28]. The residue Arg^{435} has been suggested to cause steric hindrance to SpA when binding to the IgG3-Fc[29]. The reported crystal structure of SpA-B and its IgG-Fc interaction site are depicted in **Figure 1D**. Of note, SpA binds to the same interface where the IgG Fc-Fc interactions take place to form the hexameric IgG platform required for complement activation.

Here we investigated the impact of IgG-Fc binding properties of SpA on the assembly of IgG molecules into hexamers both in solution and on antigenic surfaces. We show that SpA blocks the formation of the hexameric C1q binding platform and as a result inhibits IgG-dependent complement activation and opsonophagocytic killing (OPK) of *S. aureus*. Our data provide an important contribution to the understanding of molecular mechanisms of complement evasion, which is crucial for the intelligent design of new therapeutic strategies to tackle infectious diseases.

Results and Discussion

SpA Binds to the Fc Region of IgG1, IgG2, and IgG4 with 1:1 Stoichiometry

To investigate whether the interaction between SpA and IgG-Fc could affect IgG-dependent complement activation, we first examined the IgG-binding properties of SpA. Most *S. aureus* strains express SpA with five highly homologous Ig-binding domains (A to E; **Figure 1C**) that can bind to the Fc region of IgG. However, the single B-domain of SpA has been extensively used for structural and biochemical studies[22,30], and thus here we

tested both a soluble SpA construct containing all five domains (SpA) and a SpA construct consisting solely of the B-domain (SpA-B).

To exclude potential interactions between the SpA proteins and IgG Fab domains, we used human monoclonal IgGs that do not bind SpA via the Fab region. This was verified by comparing the binding of these antibodies to beads coated with different forms of SpA-B (**Figure S2A**). While SpA-B domain wild type (SpA-B) or SpA-B domain with abolished Fab-binding but intact Fc-binding properties (SpA-B^{AA}; D36A and D37A mutations) bound to IgG1 antibodies, SpA-B domain with abolished Fc-binding but intact Fab-binding properties (SpA-B^{KK}; Q9K and Q10K mutations) did not interact with IgG1 (**Figure S2A** and **Table S1**). As anticipated, none of the SpA-B forms bound to IgG3 antibodies (**Figure S2A**). As a reference binding experiment for all antibodies, including IgG3, we used protein G beads (protein G binds to all IgG subclasses, including IgG3) and also measured the binding of a VH3 family antibody (anti-Hla IgG1) as a control for Fab binding to SpA-B^{KK}. Of note, we further confirmed by native mass spectrometry (native MS) that the KK mutations of SpA-B^{KK} effectively impaired binding to Fc domains (**Figure S2B** and **C**). Native MS analyzes masses of intact protein complexes in a native state, allowing noncovalent interactions to remain intact[31,32].

We used enzyme-linked immunosorbent assay and native MS to study the interactions between the SpA constructs and human IgGs. In line with previous studies, our results indicate that both SpA constructs bound strongly to the Fc region of all human IgG subclasses, except IgG3 (**Figure 2** and **Figure S3** and **Table S2**). Interestingly, the binding stoichiometry differed between SpA and SpA-B. As expected with the presence of two identical binding sites on the IgG-Fc, single-domain SpA-B could—and indeed did—bind to IgG molecules with a stoichiometry of 2:1 (**Figure 2A**). SpA, however, although containing five IgG-binding domains, was found to bind principally to a single IgG molecule with a stoichiometry of 1:1 (**Figure 2B**). This suggests that full-length SpA may bind both IgG-Fc binding sites simultaneously.

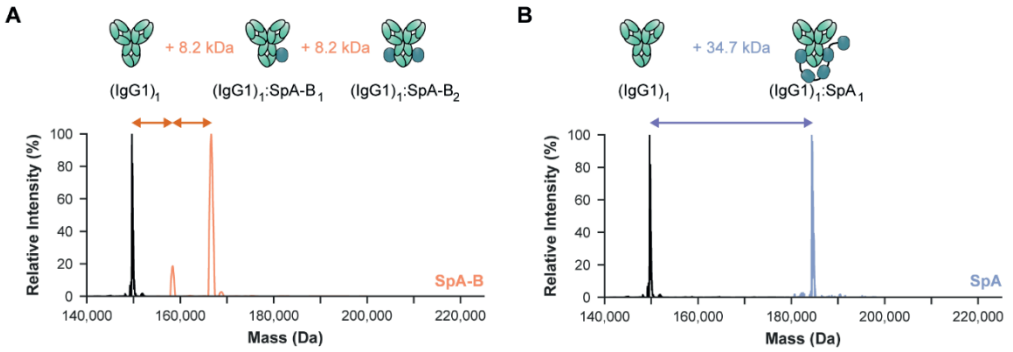


Figure 2 | Single-domain SpA-B binds IgG1 with 2:1 stoichiometry, whereas five-domain SpA binds 1:1. (A, B) Deconvoluted native mass spectra show that the mass of IgG1 (black) is shifted when incubated with SpA-B (orange) (**A**) or SpA (blue) (**B**). This shift corresponds to the binding of one or two copies of SpA-B to an IgG1 molecule and of only one SpA molecule to a single IgG1 molecule. The cartoon regarding $(IgG1)_1$:SpA_1 binding is speculative in the sense it is still unknown which SpA domain has contact with which binding site on the IgG molecules.

SpA Binding Prevents IgG Molecules from Oligomerizing

Next, we evaluated whether SpA binding could reduce IgG hexamerization in solution. Although the formation of IgG oligomers is a surface phenomenon that requires IgG binding to antigenic surfaces, this process can be mimicked in solution by introducing three mutations in the Fc region of IgG that enhance Fc-Fc interactions[7,32]. Combined mutation of residues E345R (Glu^{345}→Arg), E430G (Glu^{430}→Gly), and S440Y (Ser^{440}→Tyr) results in IgG-RGY, which readily forms hexamers in solution in a dynamic equilibrium[7,32]. Here we used native MS to investigate how SpA affects the monomer-hexamer equilibrium of IgG-RGY. As previously reported[7,32], native mass spectra of IgG1-RGY showed the presence of monomeric [denoted $(IgG1)_1$] and hexameric [$(IgG1)_6$] species, with intermediate states observed at lower abundance (**Figure 3A** and **Table S3**). For IgG subclasses that bind SpA, the relative abundance of IgG oligomers was drastically reduced on incubation with either SpA-B or SpA (**Figure 3A** and **B** and **Table S3**). Under these conditions, we observed strong binding of SpA-B/SpA to IgG monomers but not to the hexameric IgG species (**Table S3**).

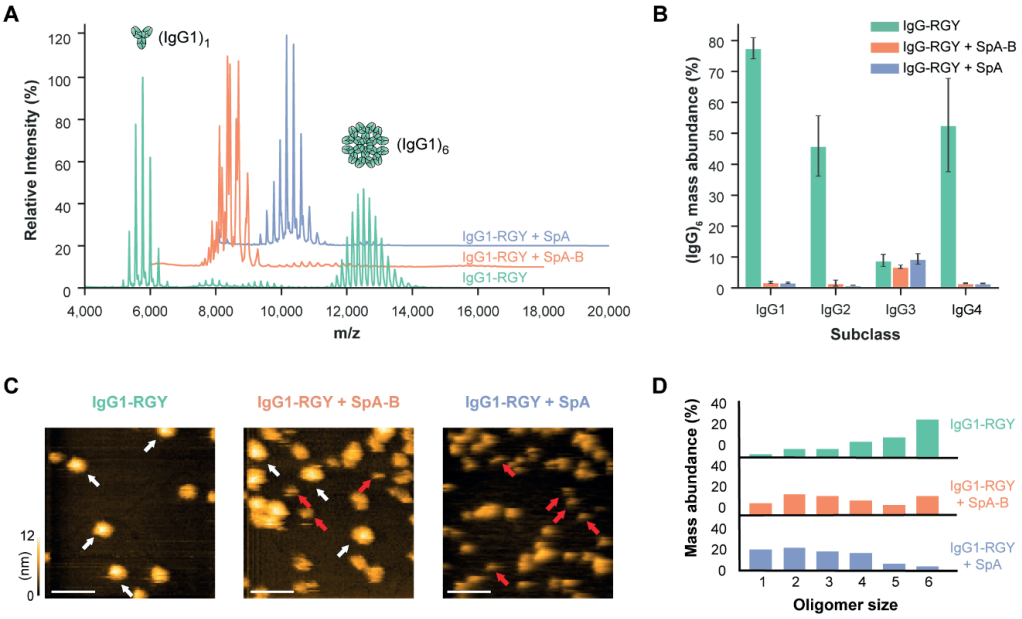


Figure 3 | Binding of SpA-B or SpA blocks IgG-RGY monomers from assembling into higher-order oligomers. (A) Native mass spectra of IgG1-RGY in the absence (green) and presence of SpA-B (orange) or SpA (blue). **(B)** Hexamer relative mass abundance of IgG-RGY subclasses in the absence and presence of SpA-B or SpA, assessed by native MS. Error bars indicate the standard deviation over three replicate samples. **(C)** HS-AFM images of IgG1-RGY on DNP-lipid bilayers in the absence and presence of SpA-B or SpA, after pre-incubation in solution. White arrows: $(IgG1)_6$. Red arrows: $(IgG1)_1$. Scale bars: 100 nm. **(D)** IgG oligomer distribution of IgG1-RGY alone (n = 372) and in the presence of SpA-B (n = 697) or SpA (n = 386) on DNP-lipid bilayers, after pre-incubation in solution. The histogram displays the fraction of IgGs constituting the respective oligomer species. Oligomer distribution was quantified by force-induced dissociation. n refers to the number of individual characterized IgGs. Statistical significance between the three experiments was evaluated with two-tailed Mann-Whitney U Tests. Control vs SpA-B: $P < 0.00001$; Control vs SpA: $P < 0.00001$; SpA-B vs SpA: $P = 0.00096$.

Native MS measurements were corroborated by high-speed atomic force microscopy (HS-AFM) experiments on 2,4-dinitrophenol (DNP)-labeled lipid-containing supported lipid bilayers (DNP-SLBs). Preincubation of the DNP-SLBs with anti-DNP antibody IgG1-RGY alone or in combination with SpA-B or SpA resulted in distinctive distributions of different-sized IgG1-RGY oligomers (**Figure 3C**). Examination of oligomer size and overall IgG1 surface density by force-induced oligomer dissociation experiments[8] allowed us to compile quantitative oligomer distributions (**Figure 3D**). Preincubation of anti-DNP IgG1-RGY with multidomain SpA led to a drastic reduction in higher-order IgG1-RGY oligomers on the DNP-SLB surface (**Figure 3D**), in agreement with our finding in the native MS experiments (**Figure 3B**). However, while the effect of both SpA constructs on the IgG1-RGY hexamer population in solution was comparable (**Figure 3B**), we observed only a ~50% reduction of IgG1-RGY hexamers when IgG1-RGY was preincubated with the SpA-B domain on DNP-SLBs (**Figure 3D**). Altogether, these data suggest that SpA binds competitively to the Fc-Fc interaction site on IgG monomers, which effectively prevents IgGs from forming higher-order IgG oligomers.

SpA Prevents the Formation of $(IgG)_6$:C1q Complexes in Solution

Having demonstrated that SpA blocks IgG hexamerization in solution, we next explored whether SpA affects the formation of $(IgG)_6$:C1q complexes. We used native MS to study the formation and behavior of these complexes in the presence and absence of SpA. In agreement with earlier data[32], native mass spectra of IgG1-RGY incubated with C1q showed clearly distinguishable species with masses corresponding to $(IgG1)_1$, $(IgG1)_6$, and $(IgG1)_6$:C1q (**Figure 4A** and **Table S4**). The experiments to study the effect of SpA were performed in two ways. When SpA or SpA-B was added to IgG1-RGY preincubated with C1q, the abundance of $(IgG1)_6$:C1q complexes was strongly reduced (**Figure 4A** and **Table S4**). Likewise, after preincubation of IgG1-RGY with the SpA constructs, the addition of C1q did not lead to detectible levels of $(IgG1)_6$:C1q complexes. Thus, regardless of the order of mixing of IgG1-RGY, C1q, and SpA, we principally detected complexes of monomeric IgG bound to SpA constructs and free unbound C1q. Similar to IgG1-RGY, both SpA constructs also prevented the assembly of $(IgG)_6$:C1q complexes for IgG2-RGY and IgG4-RGY (**Figure 4B**). Overall, these results suggest that SpA prevents the binding of C1q to IgG by blocking the formation of the hexameric IgG platform in solution.

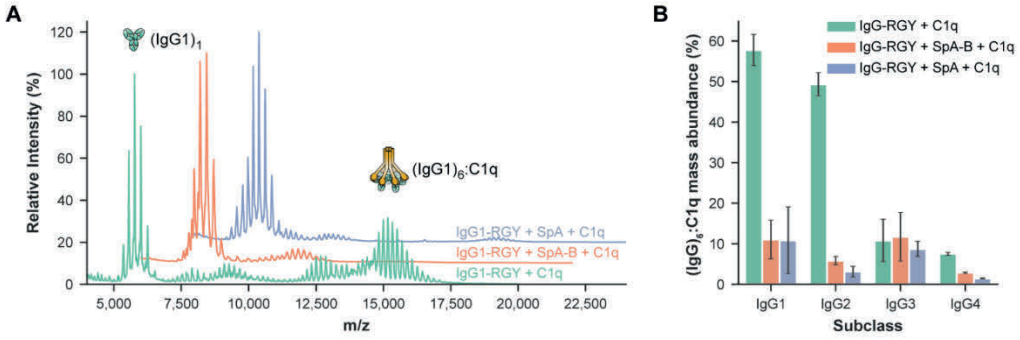


Figure 4 | SpA-B and SpA prevent the assembly $(IgG)_6$:C1q complexes in solution. (A) Native mass spectra of IgG1-RGY:C1q in the absence (green) and presence of SpA-B (orange) or SpA (blue). **(B)** The relative mass abundance of $(IgG)_6$:C1q complexes in the absence and presence SpA-B or SpA, assessed by native MS. Error bars indicate the standard deviation over three replicate samples.

SpA Inhibits Binding of C1q and C1 to Antigen-Bound IgGs on Target Surfaces

Next, we assessed how these observations for stabilized hexamers of mutant IgG-RGY in solution compare to the behavior of wild-type IgGs bound to antigenic surfaces. Since SpA and SpA-B block the formation of IgG hexamers through the same mechanism, we focused the following experiments on SpA-B. Using a quartz crystal microbalance (QCM) that we have previously used to quantitatively study the binding and oligomerization of IgG1 to antigenic SLBs[9], we assessed the association of wild-type IgG1 to similar DNP-SLBs as used for HS-AFM experiments and followed the subsequent binding of C1q in the presence or absence of SpA-B (**Figure 5A** and **B**). Monitoring the resonance frequency change of the DNP-SLB–covered SiO_2-coated quartz crystal, which is proportional to the change in bound mass, yields characteristic binding curves of anti-DNP IgG1-WT interacting with DNP-SLBs (**Figure 5A-C**, green time interval). As is evident from the constancy of the sensorgrams in the subsequent time interval (**Figure 5 A-C**; gray time interval), the

removal of IgG1-WT from the running buffer did not induce the dissociation of considerable amounts of IgG1-WT, reflecting stable binding. After having established equal amounts of surface-bound IgGs in all three experiments, we added C1q in the absence (**Figure 5A**, green time interval) and the presence of SpA-B (**Figure 5B**, orange time interval), which resulted in distinct binding curves. While C1q robustly associated with IgG1-WT on our antigenic membranes (**Figure 5A**), simultaneous incubation of C1q with SpA-B strongly reduced the binding signal (**Figure 5B**). The addition of SpA-B alone (without C1q; **Figure 5C**) resulted in a strong association of SpA-B to IgG1-WT with no detectable dissociation when SpA-B was removed from the running buffer, indicative of a high-affinity interaction. Notably, the remaining frequency shift (and thus the associated mass) at the end of the dissociation phase equaled the corresponding frequency shift for the C1q + SpA-B mixture, suggesting that the presence of SpA-B effectively precludes stable C1q binding (**Figure 5B**).

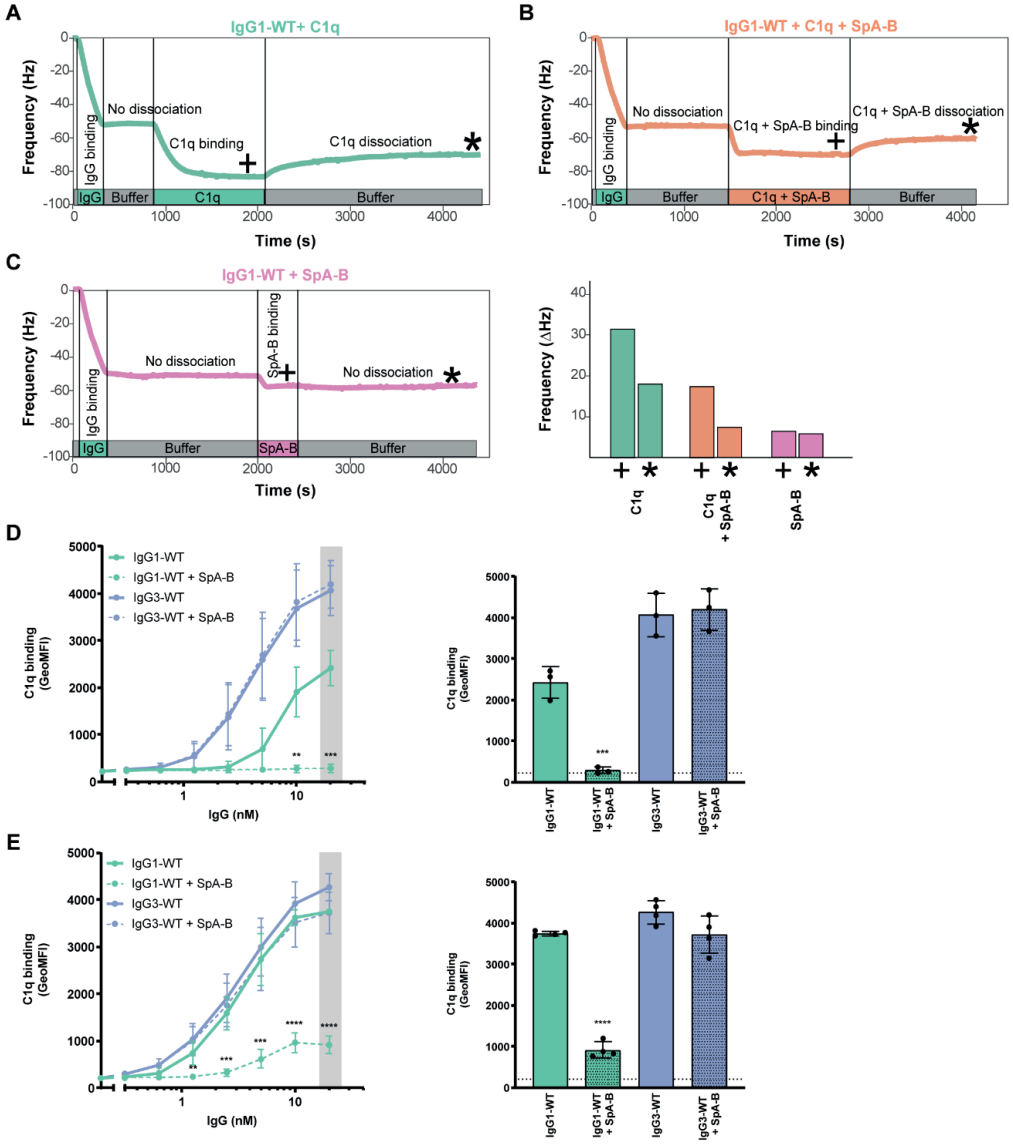


Figure 5 | SpA-B inhibits binding of C1q to antigen-bound IgGs on target surfaces. (A-C) QCM sensorgrams of C1q alone (**A**), C1q and SpA-B (**B**), or SpA-B alone (**C**) binding to anti-DNP wild-type IgG1 antibodies (IgG1-WT) bound to DNP-decorated supported lipid bilayers. The binding of C1q (14 nM) was followed in real-time in the absence or presence of SpA-B. The binding of SpA-B alone was monitored in a similar experiment without C1q. Bars represent the respective equilibrium level (+) and the level at the end of the dissociation phase (*). **(D, E)** C1q binding on IgG1-WT and IgG3-WT bound to DNP-coated beads, after incubation of C1q (**D**) or C1 complex (**E**) in the absence (solid lines) or presence (dotted lines) of SpA-B, detected with FITC-conjugated rabbit F(ab')$_2$ anti-human C1q by flow-cytometry. Bars represent the same data for the 20 nM IgG concentration only and the black dotted line shows the background fluorescence from beads that were not incubated with IgG. Data are presented as Geometric Mean Fluorescence Intensity (GeoMFI) means ± SD of three or four independent experiments. Statistical analysis was performed using an unpaired two-tailed t-test to compare buffer and SpA-B conditions and showed when significant as **P < 0.01, ***P < 0.001 and ****P < 0.0001.

QCM observations were corroborated by flow cytometry analyses, in which DNP-coated beads[33] were first labeled with anti-DNP IgG1-WT or IgG3-WT as a control. As expected, the presence of SpA-B inhibited C1q binding to IgG1-WT–labeled beads, but binding to IgG3-WT was not altered (**Figure 5D**). As a control, we confirm that binding of IgG1-WT and IgG3-WT to DNP-coated beads was similar (**Figure S4A**), while SpA-B could bind only to IgG1-WT–loaded target antigen (**Figure S4B**). Interestingly, SpA-B also blocked C1q binding when incubated after IgG1-WT:C1q complexes were formed (**Figure S4C**).

Because the classical complement pathway is initiated by C1q that is in complex with C1r and C1s proteases, we repeated the same experiments using $C1qr_2s_2$ complexes instead of C1q alone. Also, for this fully assembled C1 complex, we observed that SpA-B strongly reduced C1q binding to DNP-bound IgG1-WT, and that its presence did not affect C1q binding to IgG3-WT (**Figure 5E** and **Figure S4D**). Finally, we confirmed that similar to SpA-B, SpA also largely reduced C1q binding to surface-bound IgGs on both bead surfaces (**Figure S5A**) and lipid membranes (**Figure S5B**). Altogether, these data show that both SpA constructs efficiently block C1q binding to surface-bound antibodies of the IgG1 subclass.

SpA Inhibits C1q Binding and Complement Activation on S. aureus

We next determined whether SpA could affect antibody-dependent complement activation on the surface of *S. aureus*. To prevent cell surface SpA (and Sbi, another Ig-binding protein of *S. aureus*)[34] from binding the antibody's Fc region, we here used a strain devoid of SpA and Sbi (NewmanΔ*spa*/*sbi*). NewmanΔ*spa*/*sbi* was labeled with a human monoclonal antibody directed against wall teichoic acid (WTA)[35], a highly abundant anionic glycopolymer that is covalently anchored to the peptidoglycan layer[36]. Because the anti-WTA antibody (clone 4497) belongs to the VH3-type family[37], it is expected to bind to SpA-B or SpA via the Fab region. Nevertheless, we showed that anti-WTA IgG1 antibodies do not bind to beads coated with $SpA\text{-}B^{KK}$ (an SpA-B mutant that binds only to the IgG-Fab region) (**Figure S6A**) and further confirmed by native MS that SpA-B and SpA bind only to the Fc region of this antibody (**Figure S6B-E**).

To measure C1q binding and downstream complement activation on the bacterial surface, we incubated IgG-labeled bacteria with human serum in combination with buffer, SpA-B, or SpA. Because serum contains not only complement proteins, but also many different antibodies, including IgGs that do not bind SpA (IgG3) and antibodies that bind SpA via their Fab region (VH3-type family Igs), we used serum depleted of naturally occurring antibodies (ΔIgG/IgM serum). In this way, we could exclusively determine the effect of SpA on complement activation by monoclonal anti-WTA antibodies. In agreement with the results from experiments using DNP beads, we observed that both SpA constructs strongly reduced C1q binding on *S. aureus* labeled with anti-WTA IgG1 antibodies (**Figure 6A**). When the bacteria were labeled with anti-WTA IgG3, C1q binding was not inhibited by the presence of SpA-B or SpA (**Figure 6B**).

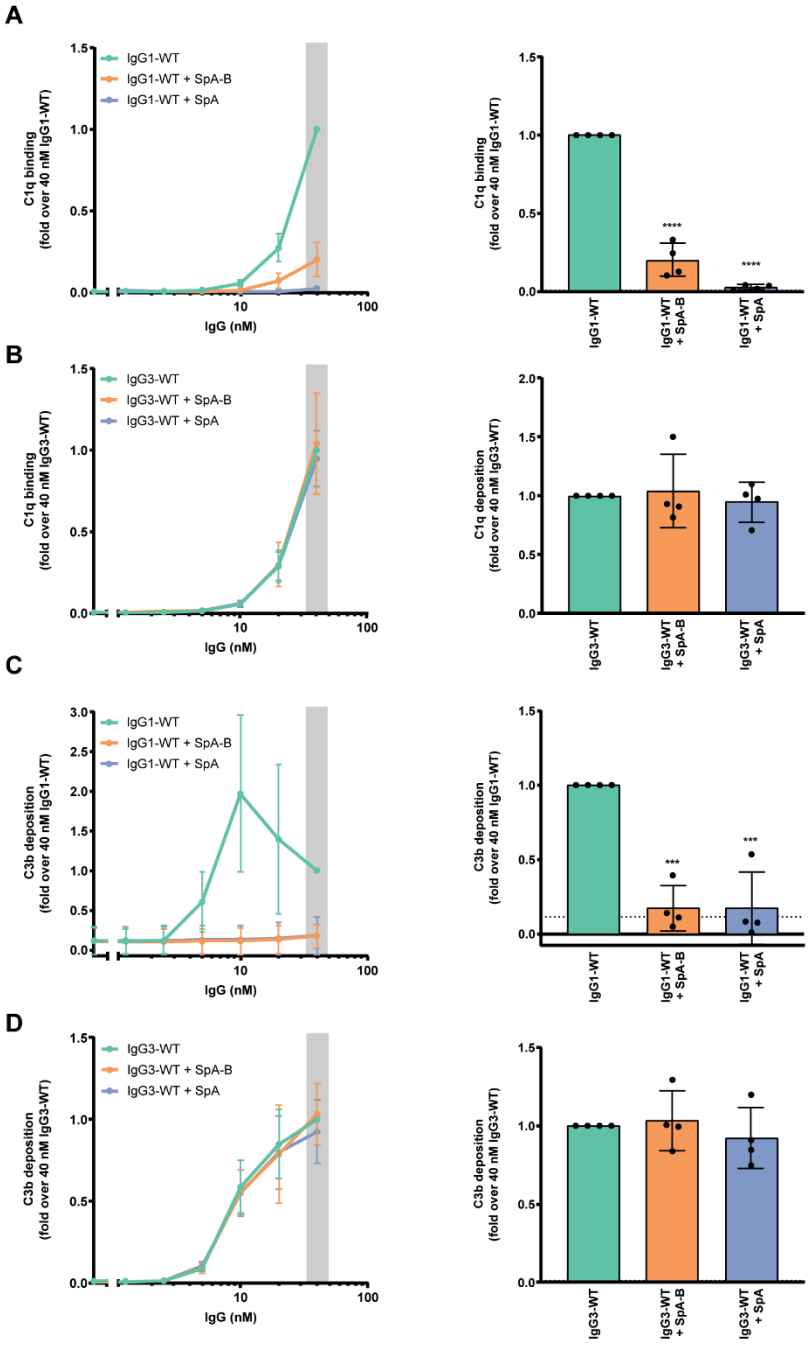


Figure 6 | SpA decreases IgG-mediated C1q binding and downstream complement on *S. aureus*. (A, B) C1q binding on anti-WTA wild-type IgG1 (IgG1-WT) (**A**) or IgG3 antibodies (IgG3-WT) (**B**) bound to NewmanΔ*spa*/*sbi* surface, after incubation of bacteria with 1% ΔIgG/IgM human serum, in the absence (green) or presence of SpA-B (orange) or SpA (blue), detected with FITC-conjugated rabbit $F(ab')_2$ anti-human C1q by flow cytometry. **(C, D)** C3b deposition on NewmanΔ*spa*/*sbi* surface after incubation of bacteria with IgG1-WT (**C**) or IgG3-WT (**D**), 1% ΔIgG/IgM human serum and buffer (green), SpA-B (orange) or SpA

(blue), detected with a monoclonal murine anti-human C3d antibody by flow cytometry. Data are presented as fold change over 40 nM concentration of IgG control ± SD of at least three independent experiments. Bars represent the same data for the 40 nM IgG concentration only and the black dotted line shows the background fluorescence from bacteria that were not incubated with IgG. Statistical analysis was performed using a one-way ANOVA to compare buffer condition with SpA-B and SpA conditions and displayed when significant as ***P < 0.001 or ****P < 0.0001.

To determine whether the inhibition of C1q binding leads to downstream inhibition of the complement cascade, we also measured the deposition of C4b and C3b. The binding of C1q to target-bound antibodies activates its attached C1r/C1s proteases that cleave C4 and C2 and generate the C3 convertase (C4b2b)[38]. Subsequently, the covalently attached C3 convertase catalyzes the deposition of C3b on the target surface[38]. On incubation of IgG1-labeled *S. aureus* with human (Ig-depleted) serum and SpA-B or SpA, surface deposition of C4b and C3b was completely abolished (**Figure S7A** and **Figure 6C**). As anticipated, C4b and C3b deposition on Newman Δ*spa*/*sbi* labeled with anti-WTA IgG3 antibodies remained unchanged in the presence of SpA-B or SpA (**Figure S7B** and **Figure 6D**). Of note, when the C3b deposition assay was repeated to include antibodies recognizing the hapten DNP as an isotype control, we did not detect C3 products on the surface of *S. aureus*. Only incubation with *S. aureus*–recognizing anti-WTA antibodies led to detectable C3b levels, confirming that this signal reflects the presence of covalently bound C3 products that were deposited on complement activation (**Figure S8**). Altogether, these findings demonstrate that SpA effectively prevents C1q recruitment and downstream activation of the complement cascade on *S. aureus*.

SpA Decreases OPK of S. aureus by IgG1 but Not IgG3

Finally, we evaluated whether the inhibitory effect of SpA on complement activation could result in less killing of *S. aureus* by neutrophils. Neutrophils are the first cells to be recruited from the blood to the site of infection, where they engulf and internalize bacteria via phagocytosis and subsequently kill them by exposure to antimicrobial agents such as antimicrobial peptides, reactive oxygen species and enzymes[39]. Previous work from our group demonstrated that the deposition of C3-derived opsonins greatly enhances the phagocytic uptake of *S. aureus*[40], because C3b molecules can be recognized by complement receptors on phagocytes. To compare the OPK activity of anti-WTA IgG1 and IgG3 antibodies, we used a wild-type *S. aureus* strain (Newman WT) that expresses both cell-wall anchored SpA and cell-associated Sbi. The Newman WT were incubated with IgG1 or IgG3 and 1% ΔIgG/IgM serum as a complement source and then mixed with freshly isolated human neutrophils to enable phagocytosis and killing. We found that IgG3 antibodies more potently induced Newman WT killing than IgG1 antibodies (**Figure 7A**).

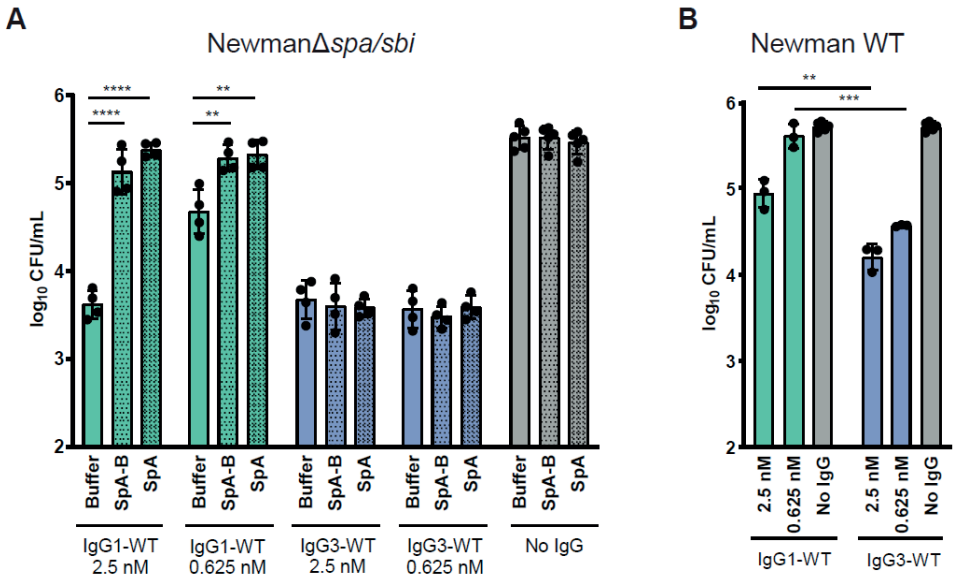


Figure 7 | SpA decreases opsonophagocytic killing of *S. aureus* by IgG1 but not IgG3. (A) CFU enumeration of NewmanΔ*spa/sbi* after incubation with anti-WTA IgG1-WT(green), IgG3-WT (blue) or no IgG (grey) in the presence of 1% ΔIgG/IgM human serum and buffer, SpA-B or SpA, followed by incubation with human neutrophils. **(B)** CFU enumeration of Newman WT after incubation with IgG1-WT (green), IgG3-WT (blue) or no IgG (grey) in the presence of 1% ΔIgG/IgM human serum, followed by incubation with human neutrophils. Data are presented as $\log_{10}$ CFU/mL ± SD of at least three independent experiments. Statistical analysis was performed using a one-way ANOVA to compare the buffer condition with SpA-B and SpA conditions (**A**) or using an unpaired two-tailed t-test to compare IgG1-WT and IgG3-WT conditions (**B**). Statistical differences were shown when significant as **P < 0.01, ***P < 0.001 or ****P < 0.0001.

To determine whether this is indeed caused by the interaction of SpA with IgG1 antibodies, we also performed the killing of NewmanΔ*spa/sbi* in the absence or presence of exogenous SpA-B or SpA. While anti-WTA IgG1 antibodies induced OPK of *S. aureus* (**Figure 7B**), we observed that the addition of soluble SpA-B or SpA blocked OPK (**Figure 7B**). As expected, killing of *S. aureus* in the presence of IgG3 remained unaffected in the presence of SpA (**Figure 7B**). Overall, these data suggest that both soluble and cell-exposed SpA reduce the OPK of *S. aureus* via IgG1 but not via IgG3.

Discussion

Antibody-dependent complement activation is an important immunologic mechanism to accelerate bacterial killing[1]. To effectively trigger complement, antibodies should bind bacterial cells and subsequently form oligomeric IgG clusters that recruit C1[7,8]. In this paper, we identify SpA from *S. aureus* as an example of a bacterial immune evasion molecule that specifically blocks antibody clustering by inhibiting IgG Fc-Fc contacts. These findings are of relevance to the basic pathophysiology of staphylococcal infections,

and also to the development of immune therapies against *S. aureus*. Furthermore, SpA could be used as a tool to better understand the role of IgG clustering in various disease processes in which antibodies and complement are involved.

Our study demonstrates that by binding to the Fc region of IgG, soluble SpA blocks IgG hexamerization, resulting in inhibition of C1q recruitment, downstream complement activation on the *S. aureus* surface, and bacterial killing by human phagocytes. Although the interference of SpA with the classical complement pathway has been noted previously, the exact molecular mechanism has remained elusive and there has been debate about its role as an activator or inhibitor[41-45] of the complement system. Thanks to recent insights into IgG oligomerization[7] and more advanced methods for directly visualizing this process, we have been able to unravel the mechanism of SpA-dependent complement inhibition in highly purified conditions. The binding of SpA to the Fc region of monomeric, but not hexameric, IgG-RGY species suggests that SpA binds exclusively to free, unoccupied Fc regions. Thus, by binding to monomeric IgGs, SpA likely prevents the transition to the hexameric state.

Since *S. aureus* strains produce SpA with four or five Ig-binding domains[46], we compared the complement inhibitory activity of SpA with five Ig-binding domains versus a single SpA-B domain. AFM experiments suggest that SpA more effectively blocks hexamerization of IgG molecules on antigenic membranes than SpA-B. Because multidomain SpA is likely to bind IgG bivalently, it dissociates at a lower rate than monovalently bound SpA-B. The resulting binding affinity advantage of SpA over SpA-B allows it to compete more effectively with the IgG Fc-Fc interactions. Indeed, multidomain SpA was more effective than SpA-B in inhibiting C1q binding to target-bound IgGs, which is consistent with the notion that multivalent SpA molecules have a complement inhibitory advantage over a single SpA domain. However, on measuring downstream complement activation, the inhibitory advantage of SpA over a single SpA domain was less evident. We speculate that the requirement for multiple domains may be even more important for complement inhibition by cell-anchored SpA. Since the majority of SpA produced by *S. aureus* is anchored to the bacterial cell wall, multiple SpA domains may be needed to provide the molecule with sufficient length and flexibility to bind Fc domains of bacterium-bound IgG. Although the majority of SpA is cell-anchored, 6.5 to 15% is secreted before sorting[20] or released from the cell wall after enzymatic cleavage by LytM[19]. We observed that the concentration of recombinant multidomain SpA necessary to decrease C1q deposition on an antigenic surface is lower than the amount reportedly secreted by Newman or USA300 strains in vitro[47].

Besides SpA, there are other bacterial proteins that bind to Ig[10]. The best known are protein G (from group C and G streptococci), M and M-like proteins (group A streptococci), protein L (*Peptostreptococcus magnus*), Sbi, and SSL10 (*S. aureus*). Protein G, protein M, M-like proteins, and Sbi all bind the Fc domain of IgGs in the vicinity of the SpA-binding site[48-50], suggesting that these IgG-Fc–binding proteins may be able to block IgG hexamerization as well.

Overall, our data provide insights crucial for the development of effective immune therapies against *S. aureus*. It is well recognized that both neutrophils and complement

play essential roles in the killing of *S. aureus*. While neutrophils can engulf *S. aureus* directly, previous work from our group demonstrated that the decoration of bacteria with C3-derived opsonins strongly enhances effective phagocytosis[40]. Therefore, the identification of antibacterial antibodies with strong complement-activating potential provides an interesting approach to boosting the host immune system and preventing or treating these infections. From this study, it is now clear that such approaches can be more successful if we take the SpA-dependent antibody modulation into account. Although most active and passive immunization approaches to develop or induce antibodies targeting *S. aureus* surface components (*e.g.*, capsule polysaccharide, lipoteichoic acid, various surface adhesins) have failed at the clinical trial stage[5], we propose that such strategies could be more effective when also blocking the effect of SpA. In this context, although our comparison of IgG1's and IgG3's potency in enabling OPK of *S. aureus* does not allow discrimination of the inhibitory effect of SpA on FcγR-mediated from complement-mediated OPK, it suggests that monoclonal antibodies targeting surface components of *S. aureus* should be developed as IgG3 (or variants thereof) that are not targeted by SpA. Alternatively, any amino acid modification in IgG1 or IgG2 subclasses that prevents SpA binding to IgG, but not IgG hexamerization, would be of value. We also propose that the use of SpA as a vaccine antigen or of monoclonal antibodies directed against Fc-binding domains of SpA might be needed to prevent the anticomplement effect of SpA and thereby increase the chance of bacterial clearance.

Furthermore, SpA could also be used as a research tool to specifically examine the role of IgG hexamerization in various disease processes. For instance, SpA or a single domain of SpA (mutated to bind only the IgG-Fc region) could be used to study whether antibodies induced during an infection require Fc-Fc contacts to induce complement activation on the invading pathogen. In fact, our data with IgG1 antibodies against WTA suggest that naturally induced antibodies against *S. aureus* indeed require Fc-Fc contacts to induce complement activation on the surface of bacteria. Although we used monoclonal antibodies, it is well recognized that IgG1 antibodies against WTA are produced during an *S. aureus* infection in vivo[51-54]. Also, given that excessive activation of complement has been associated with the clinical manifestation of several autoimmune diseases, SpA could be used to study whether autoreactive antibodies induce IgG clustering on altered host cells. Finally, SpA, or molecules targeting the SpA-binding site in IgG, can potentially block unwanted antibody responses. Indeed, the therapeutic potential of SpA has been tested for the treatment of autoimmune disorders[55-57], although the rationale for using SpA was based on its effectiveness as a B cell superantigen.

In conclusion, the identification of SpA as a biological inhibitor of IgG hexamerization will increase our understanding of antibody-dependent immunologic mechanisms and may help accelerate the development of immune interventions in infection and inflammation.

Materials and Methods

More details on the materials and methods used in this study are provided in the **Supplementary Materials and Methods**.

Native MS

Native MS analyses were performed on a modified LCT time-of-flight instrument (Waters) or a standard Exactive Plus EMR Orbitrap instrument (Thermo Fisher Scientific). Before analysis, buffers were exchanged to 150 mM ammonium acetate (pH 7.5) through six consecutive dilution and concentration steps at 4 °C using Amicon Ultra centrifugal filters with a 3-kDa or 10-kDa molecular weight cutoff (Merck). Protein complexes were assembled by mixing the subcomponents at the desired molar ratios, followed by incubation at room temperature (RT) for at least 30 min. For experiments studying the effect of SpA constructs, the incubation step with SpA proceeded for at least 3 h at 37 °C due to the relatively slow disassembly rate of the IgG-RGY hexamers. IgG-RGY hexamers were measured at a total IgG concentration of 2 μM in the presence or absence of 10 μM SpA-B or SpA (ProSpec). For measurements of $(IgG)_6$:C1q complexes, 0.5 μM C1q (Complement Technology) was used. Samples were loaded into gold-coated borosilicate capillaries (prepared in house) for direct infusion from a static nano-electrospray ionization source. Relative abundances of protein complexes were determined using an in-house script that sums and compares ion intensities of the different species, similar to a previously described method[58]. Deconvoluted mass spectra were generated by Bayesian deconvolution using UniDec[59].

HS-AFM

HS-AFM (Research Institute of Biomolecule Metrology) was conducted in tapping mode at RT in buffer, with free amplitudes of 1.5 to 2.5 nm and amplitude set points > 90%. Silicon nitride cantilevers with electron beam-deposited tips (USC-F1.2-k0.15; Nanoworld AG), nominal spring constants of 0.15 N m^{-1}, resonance frequencies around 500 kHz, and a quality factor of ~2 in liquids were used. Imaging was performed in buffer 1 (10 mM Hepes, 150 mM NaCl, and 2 mM $CaCl_2$, pH 7.4). All IgGs were diluted and incubated in the same buffer.

DNP-labeled SLBs for HS-AFM were prepared on muscovite mica. The liposomes were incubated on the freshly cleaved surface (500 μg mL^{-1} in buffer 1), placed in a humidity chamber to prevent evaporation, and heated to 60 °C for 30 min. Then the temperature was gradually cooled to RT within 30 min, followed by exchanging the solution with buffer 1. After 10 min of equilibration at RT and 15 more buffer exchanges, the SLB was ready for imaging. In order to passivate any exposed mica, SLBs were incubated with 333 nM IgG1-b12 (irrelevant human IgG1 control antibody against HIV-1 gp120)[60] for 10 min before the molecules of interest were added.

QCM

QCM experiments were done using a two-channel QCM-I system (MicroVacuum). AT cut SiO_2-coated quartz crystals with a diameter of 14.0 mm and a resonance frequency of 5 MHz were used (Quartz Pro AB). All sensorgrams were recorded on the first, third, and fifth harmonic frequencies. The data shown are related to the third harmonic. Before each set of experiments, the SiO_2-coated crystals were cleaned by immersion in 2% sodium

dodecyl sulfate (SDS) for 30 min, followed by thorough rinsing with Milli-Q H_2O. The chips were dried in a gentle stream of N_2 and oxidized using air plasma (4 min at 80 W), then mounted in the measurement chamber. The sensor surface was further cleaned by a flow of 2% SDS at 250 µL min^{-1} for 5 min, followed by Milli-Q H_2O at 250 µL min^{-1} for 5 min directly before the measurements. Finally, IgG-free buffer was injected for equilibration at 50 µL min^{-1}. All subsequent injections were performed at 50 µL min^{-1}. To generate DNP-SLBs on QCM chips, the DPPC:DPPE:DNP-cap-DPPE liposome stock solution was heated to 60 °C for 30 min and then slowly cooled to RT within 30 min. The solution was ready for injection after dilution to 200 µg/mL with buffer 1. DNP-SLB formation was typically complete after 30 min at 50 µL min^{-1}, after which the flow medium was changed to buffer 1 for equilibration.

DNP-Coated Beads Assays

Streptavidin beads (Dynabeads M-270; Invitrogen) were washed in PBS-TH (phosphate-buffered saline [PBS], 0.05% [vol/vol] Tween-20, and 0.5% human serum albumin [HSA]) and incubated (diluted 100×) with 1 µg/mL biotinylated DNP (DNP-PEG2-GSGSGSGK(Biotin)-NH2; Pepscan Therapeutics) in PBS-TH for 30 min at 4 °C with shaking. For each condition, 0.5 µL of beads was used (~3×10^5 beads/condition). After two washes with PBS-TH, DNP-coated beads were incubated with 20 nM anti-DNP IgG or twofold serial dilutions of anti-DNP IgG (starting from 20 nM IgG) for 30 min at 4 °C with shaking (±700 rpm). The following incubation steps were performed under shaking conditions (±700 rpm) for 30 min at 4 °C unless stated otherwise. IgG-bound DNP-coated beads were washed twice with VBS-TH (Veronal buffered saline [pH 7.4], 0.5 mM $CaCl_2$, 0.25 mM $MgCl_2$, 0.05% [vol/vol] Tween-20, and 0.5% HSA) and incubated with 1.3 nM of purified C1q or C1 (Complement Technology) alone, in combination with, or followed by incubation of 200 nM or fourfold dilutions (starting from 1 µM) of recombinant SpA-B or SpA (ProSpec) in VBS-TH at 37 °C. Finally, the beads were washed twice with PBS-TH and incubated with 4 µg/mL fluorescein isothiocyanate (FITC)-conjugated rabbit $F(ab')_2$ anti-human C1q. C1q binding to the beads was detected using flow cytometry (BD FACSVerse), and data were analyzed based on a single bead population using FlowJo software.

Complement Deposition Assays on S aureus Surface

S. aureus NewmanΔ*spa*/*sbi* strain was fluorescently labeled by transformation with the pCM29 plasmid, constitutively expressing mAmetrine under regulation of the sarA promotor, as described previously[61,62]. Bacteria were grown overnight in Todd Hewitt broth (THB) plus 10 µg/mL chloramphenicol, diluted to an OD_{600} of 0.05 in fresh THB plus chloramphenicol, and cultured until midlog phase (OD_{600} = 0.5). Cells were collected, washed, resuspended in RPMI-H medium (RPMI + 0.05% HSA), and aliquoted at −20 °C.

Similar to the Dynabeads assays, all incubation steps were performed under shaking conditions (±700 rpm) for 30 min at 4 °C (unless stated otherwise), followed by a single wash with RPMI-H by centrifugation. Bacteria (7.5×10^5 CFU) were incubated with twofold titration (starting from 40 nM) IgG in RPMI-H, followed by incubation with 1% ΔIgG/IgM serum in combination with buffer and 200 nM of SpA-B or SpA in RPMI-H for 30 min at 37 °C. For C1q detection, bacteria were incubated with 0.5 µg/mL chicken anti-human C1qA (Sigma-Aldrich), followed by incubation with phycoerythrin-conjugated donkey $F(ab')_2$ anti-chicken (Jackson ImmunoResearch), diluted 1:500 in RPMI-H. For C4b

and C3b detection, 1 μg/mL murine anti-C4d (Quidel) or anti-C3d (Quidel) antibody was incubated with bacteria, respectively. Subsequently, bacteria were incubated with FITC-conjugated goat $F(ab')_2$ anti-mouse (Dako), diluted 1:100 in RPMI-H. After labeling, samples were fixed with 1% paraformaldehyde in RPMI-H, and the binding of C1q, C4b, and C3b to bacteria was detected using flow cytometry (BD FACSVerse). Data were analyzed using FlowJo software.

OPK of S aureus by Neutrophils

Human neutrophils were purified from blood of healthy donors by the Ficoll/Histopaque density gradient method[63]. *S. aureus* Newman WT and Newman*Δspa*/*sbi* constitutively expressing mAmetrine were freshly grown to midlog phase, washed, and opsonized as follows. Newman WT were incubated with fourfold titration (starting from 10 nM) of anti-WTA IgG1 or IgG3 and 1% ΔIgG/IgM serum in Hank's balanced salt solution (HBSS) + 0.1% HSA (HBSS-H). Newman*Δspa*/*sbi* were incubated with 2.5 nM anti-WTA IgG1 or IgG3 and 1% ΔIgG/IgM serum in the absence or presence of exogenous SpA-B or SpA (200 nM) in HBSS-H. After 30 min at 37 °C, the bacteria (8.5×10^5 CFU) were incubated with freshly isolated neutrophils for 90 min under 5% CO_2 at 37 °C, at a 1:1 bacteria:cell ratio. Subsequently, neutrophils were lysed with cold 0.3% (wt/vol) saponin in water for up to 15 min on ice, and samples were serially diluted in PBS and plated in duplicate onto TSA plates. The plates were incubated overnight at 37 °C, and viable bacteria were quantified by CFU enumeration.

Ethical Statement

Human serum and blood were obtained from healthy donors after informed consent was obtained from all subjects, in accordance with the Declaration of Helsinki. Approval from the Medical Ethics Committee of the University Medical Center Utrecht was obtained (METC protocol 07-125/C, approved March 1, 2010).

Data Availability

All study data are included in the main text and Supporting Information.

Acknowledgements

We thank Annette Stemerding for fruitful discussions. This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs (TTW-NACTAR Grant #16442 [to AJRH and SHMR]). This work was supported by the European Union's Horizon 2020 research program H2020-MSCA-ITN (675106, to J.A.G.v.S. and F.B.), and a European Research Council (ERC) Starting Grant (639209, to S.H.M.R.). M.A.d.B. and A.J.R.H. further acknowledge funding for the large-scale proteomics facility, the Netherlands Proteomics Center, through the X-omics Road Map program (Project

184.034.019) and the EU Horizon 2020 program Epic-XS (Project 823839). A.J.R.H. and S.H.M.R. acknowledge the Utrecht University Molecular Immunology Hub. J.P. receives support from the European Fund for Regional Development (EFRE, IWB2020), the Federal State of Upper Austria, and the Austrian Science Fund (FWF, P33958 and P34164).

References

1. Lu, L.L., Suscovich, T.J., Fortune, S.M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nature Reviews Immunology* **18**, 46-61 (2018).
2. Lee, C.-H. et al. IgG Fc domains that bind C1q but not effector Fcγ receptors delineate the importance of complement-mediated effector functions. *Nat. Immunol.* **18**, 889-898 (2017).
3. Laxminarayan, R. et al. Antibiotic resistance—the need for global solutions. *The Lancet Infectious Diseases* **13**, 1057-1098 (2013).
4. Theuretzbacher, U. & Piddock, L.J.V. Non-traditional Antibacterial Therapeutic Options and Challenges. *Cell Host Microbe* **26**, 61-72 (2019).
5. Sause, W.E., Buckley, P.T., Strohl, W.R., Lynch, A.S. & Torres, V.J. Antibody-Based Biologics and Their Promise to Combat Staphylococcus aureus Infections. *Trends Pharmacol. Sci.* **37**, 231-241 (2016).
6. Gaboriaud, C. et al. Structure and activation of the C1 complex of complement: unraveling the puzzle. *Trends Immunol.* **25**, 368-373 (2004).
7. Diebolder, C.A. et al. Complement Is Activated by IgG Hexamers Assembled at the Cell Surface. *Science* **343**, 1260-1263 (2014).
8. Strasser, J. et al. Unraveling the Macromolecular Pathways of IgG Oligomerization and Complement Activation on Antigenic Surfaces. *Nano Lett.* **19**, 4787-4796 (2019).
9. Strasser, J. et al. Weak Fragment Crystallizable (Fc) Domain Interactions Drive the Dynamic Assembly of IgG Oligomers upon Antigen Recognition. *ACS Nano* **14**, 2739-2750 (2020).
10. Sidorin, E.V. & Solov'eva, T.F. IgG-binding proteins of bacteria. *Biochemistry (Moscow)* **76**, 295-308 (2011).
11. Lowy, F.D. Staphylococcus aureus Infections. *New England Journal of Medicine* **339**, 520-532 (1998).
12. Kim, H.K., Kim, H.-Y., Schneewind, O. & Missiakas, D. Identifying protective antigens of Staphylococcus aureus, a pathogen that suppresses host immune responses. *The FASEB Journal* **25**, 3605-3612 (2011).
13. Palmqvist, N., Foster, T., Tarkowski, A. & Josefsson, E. Protein A is a virulence factor in Staphylococcus aureus arthritis and septic death. *Microb. Pathog.* **33**, 239-249 (2002).
14. Falugi, F., Kim, H.K., Missiakas, D.M. & Schneewind, O. Role of Protein A in the Evasion of Host Adaptive Immune Responses by Staphylococcus aureus. *mBio* **4**, e00575-00513 (2013).
15. Kim, H.K., Cheng, A.G., Kim, H.-Y., Missiakas, D.M. & Schneewind, O. Nontoxigenic protein A vaccine for methicillin-resistant Staphylococcus aureus infections in mice. *J. Exp. Med.* **207**, 1863-1870 (2010).
16. Gatlin, C.L. et al. Proteomic profiling of cell envelope-associated proteins from Staphylococcus aureus. *Proteomics* **6**, 1530-1549 (2006).
17. Ventura, C.L. et al. Identification of a novel Staphylococcus aureus two-component leukotoxin using cell surface proteomics. *PLoS One* **5**, e11634 (2010).
18. Ythier, M. et al. Proteomic and Transcriptomic Profiling of Staphylococcus aureus Surface LPXTG-proteins: Correlation with agr Genotypes and Adherence Phenotypes*. *Mol. Cell. Proteomics* **11**, 1123-1139 (2012).
19. Becker, S., Frankel, M.B., Schneewind, O. & Missiakas, D. Release of protein A from the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci.* **111**, 1574-1579 (2014).
20. O'Halloran, D.P., Wynne, K. & Geoghegan, J.A. Protein A Is Released into the Staphylococcus aureus Culture Supernatant with an Unprocessed Sorting Signal. *Infection and Immunity* **83**, 1598-1609 (2015).
21. Uhlén, M. et al. Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. *J. Biol. Chem.* **259**, 1695-1702 (1984).

22. Deisenhofer, J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-.ANG. resolution. *Biochemistry* **20**, 2361-2370 (1981).
23. Sasso, E.H., Silverman, G.J. & Mannik, M. Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. *J. Immunol.* **142**, 2778-2783 (1989).
24. Sasso, E.H., Silverman, G.J. & Mannik, M. Human IgA and IgG F(ab')2 that bind to staphylococcal protein A belong to the VHIII subgroup. *J. Immunol.* **147**, 1877-1883 (1991).
25. Graille, M. et al. Crystal structure of a *Staphylococcus aureus* protein A domain complexed with the Fab fragment of a human IgM antibody: Structural basis for recognition of B-cell receptors and superantigen activity. *Proc. Natl. Acad. Sci.* **97**, 5399-5404 (2000).
26. Goodyear , C.S. & Silverman , G.J. Death by a B Cell Superantigen : In Vivo VH-targeted Apoptotic Supraclonal B Cell Deletion by a Staphylococcal Toxin. *J. Exp. Med.* **197**, 1125-1139 (2003).
27. Deis, Lindsay N. et al. Multiscale Conformational Heterogeneity in Staphylococcal Protein A: Possible Determinant of Functional Plasticity. *Structure* **22**, 1467-1477 (2014).
28. Jendeberg, L. et al. Engineering of Fc1 and Fc3 from human immunoglobulin G to analyse subclass specificity for staphylococcal protein A1In this paper, amino acid residues in Fc are numbered, as in Deisenhofer (1981), using the Eu numbering (Kabat et al., 1991).1. *J. Immunol. Methods* **201**, 25-34 (1997).
29. Shah, I.S., Lovell, S., Mehzabeen, N., Battaile, K.P. & Tolbert, T.J. Structural characterization of the Man5 glycoform of human IgG3 Fc. *Mol. Immunol.* **92**, 28-37 (2017).
30. Gouda, H. et al. NMR Study of the Interaction between the B Domain of Staphylococcal Protein A and the Fc Portion of Immunoglobulin G. *Biochemistry* **37**, 129-136 (1998).
31. Leney, A.C. & Heck, A.J. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass. Spectrom.* **28**, 5-13 (2017).
32. Wang, G. et al. Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Mol. Cell* **63**, 135-145 (2016).
33. Zwarthoff, S.A. et al. Functional Characterization of Alternative and Classical Pathway C3/C5 Convertase Activity and Inhibition Using Purified Models. *Front. Immunol.* **9** (2018).
34. Zhang, L., Jacobsson, K., Vasi, J., Lindberg, M. & Frykberg, L. A second IgG-binding protein in Staphylococcus aureus. *Microbiology* **144**, 985-991 (1998).
35. Lehar, S.M. et al. Novel antibody–antibiotic conjugate eliminates intracellular S. aureus. *Nature* **527**, 323-328 (2015).
36. Brown, S., Jr., J.P.S.M. & Walker, S. Wall Teichoic Acids of Gram-Positive Bacteria. *Annu. Rev. Microbiol.* **67**, 313-336 (2013).
37. Fong, R. et al. Structural investigation of human S. aureus-targeting antibodies that bind wall teichoic acid. *MAbs* **10**, 979-991 (2018).
38. Law, S.K.A. & Dodds, A.W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6**, 263-274 (1997).
39. Nathan, C. Neutrophils and immunity: challenges and opportunities. *Nature Reviews Immunology* **6**, 173-182 (2006).
40. Rooijakkers, S.H.M. et al. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nat. Immunol.* **6**, 920-927 (2005).
41. Sjöquist, J. & Stålenheim, G. Protein A from Staphylococcus aureus. IX. Complement-fixing activity of protein A-IgG complexes. *J. Immunol.* **103**, 467-473 (1969).
42. Kronvall, G. & Gewurz, H. Activation and inhibition of IgG mediated complement fixation by staphylococcal protein A. *Clin Exp Immunol* **7**, 211-220 (1970).
43. Stålenheim, G. & Castensson, S. Protein A from Staphylococcus aureus conversion of complement factor C3 by aggregates between IgG and protein A. *FEBS Lett.* **14**, 79-81 (1971).
44. Stålenheim, G. Protein A from Staphylococcus aureus XI. Fixation of human complement and of complement from guinea pig and rabbit. *Acta Pathol. Microbiol. Scand. B Microbiol. Immunol.* **79**, 665-672 (1971).
45. Stålenheim, G., Götze, O., Cooper, N.R., Sjöquist, J. & Müller-Eberhard, H.J. Consumption of human complement components by complexes of IgG with protein A of Staphylococcus aureus. *Immunochemistry* **10**, 501-507 (1973).
46. Baum, C. et al. Non-spa-Typeable Clinical Staphylococcus aureus Strains Are Naturally Occurring Protein A Mutants. *J. Clin. Microbiol.* **47**, 3624-3629 (2009).
47. Hoppenbrouwers, T. et al. Staphylococcal Protein A Is a Key Factor in Neutrophil Extracellular Traps Formation. *Front. Immunol.* **9** (2018).

48. Atkins, K.L. et al. S. aureus IgG-binding proteins SpA and Sbi: Host specificity and mechanisms of immune complex formation. *Mol. Immunol.* **45**, 1600-1611 (2008).
49. Sauer-Eriksson, A.E., Kleywegt, G.J., Uhlén, M. & Jones, T.A. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3**, 265-278 (1995).
50. Frick, I.M. et al. Convergent evolution among immunoglobulin G-binding bacterial proteins. *Proc. Natl. Acad. Sci.* **89**, 8532-8536 (1992).
51. Jung, D.-J. et al. Specific Serum Ig Recognizing Staphylococcal Wall Teichoic Acid Induces Complement-Mediated Opsonophagocytosis against Staphylococcus aureus. *J. Immunol.* **189**, 4951-4959 (2012).
52. Lee, J.-H. et al. Surface Glycopolymers Are Crucial for *In Vitro* Anti-Wall Teichoic Acid IgG-Mediated Complement Activation and Opsonophagocytosis of Staphylococcus aureus. *Infection and Immunity* **83**, 4247-4255 (2015).
53. Kurokawa, K., Takahashi, K. & Lee, B.L. The staphylococcal surface-glycopolymer wall teichoic acid (WTA) is crucial for complement activation and immunological defense against Staphylococcus aureus infection. *Immunobiology* **221**, 1091-1101 (2016).
54. van Dalen, R. et al. Do not discard Staphylococcus aureus WTA as a vaccine antigen. *Nature* **572**, E1-E2 (2019).
55. Eftimiadi, G., Vinai, P. & Eftimiadi, C. Staphylococcal protein A as a pharmacological treatment for autoimmune disorders. *J Autoimmune Disord* **3**, 40 (2017).
56. Bernton, E., Gannon, W., Kramer, W. & Kranz, E. PRTX-100 and methotrexate in patients with active rheumatoid arthritis: A Phase Ib randomized, double-blind, placebo-controlled, dose-escalation study. *Clinical Pharmacology in Drug Development* **3**, 477-486 (2014).
57. Viau, M. & Zouali, M. Effect of the B cell superantigen protein A from S. aureus on the early lupus disease of (NZB×NZW) F1 mice. *Mol. Immunol.* **42**, 849-855 (2005).
58. Wang, G., Johnson, A.J. & Kaltashov, I.A. Evaluation of Electrospray Ionization Mass Spectrometry as a Tool for Characterization of Small Soluble Protein Aggregates. *Anal. Chem.* **84**, 1718-1724 (2012).
59. Marty, M.T. et al. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370-4376 (2015).
60. Burton, D.R. et al. Efficient Neutralization of Primary Isolates of HIV-1 by a Recombinant Human Monoclonal Antibody. *Science* **266**, 1024-1027 (1994).
61. Pang, Y.Y. et al. agr-Dependent Interactions of Staphylococcus aureus USA300 with Human Polymorphonuclear Neutrophils. *J. Innate Immun.* **2**, 546-559 (2010).
62. Schenk, S. & Laddaga, R.A. Improved method for electroporation of Staphylococcus aureus. *FEMS Microbiol. Lett.* **94**, 133-138 (1992).
63. Surewaard, B.G., van Strijp, J.A. & Nijland, R. Studying interactions of Staphylococcus aureus with neutrophils by flow cytometry and time lapse microscopy. *J Vis Exp*, e50788 (2013).

Supporting Information

Supplementary Methods

Production and purification of human monoclonal antibodies

Monoclonal antibodies against DNP (DNP-G2a2) used for MS, HS-AFM and QCM experiments were recombinantly expressed as wild-type and hexamer-forming RGY mutant[1] IgG1, IgG2, IgG3 and IgG4 and obtained from Genmab (Utrecht, the Netherlands)[2,3].

For the other experiments of this study, we used human monoclonal antibodies produced recombinantly in human Expi293F cells (Life Technologies) as described before[4], with minor modifications. Briefly, gBlocks (Integrated DNA technologies, IDT), containing codon-optimized variable heavy and light chain (VH and VL) sequences with an upstream KOZAK and HAVT20 signal peptide, were cloned into homemade pcDNA34 vectors, upstream the IgG heavy and kappa light chain constant regions, respectively, using Gibson assembly (New England Biolabs). The VH and VL sequences of the antibodies were derived from previously reported antibodies anti-DNP (DNP-G2a2)[5] and anti-WTA GlcNAc-β-4497 (patent WO/2014/193722)[6] (**Table S5**). Transfection of EXPI293F cells was performed using PEI (Polyethylenimine HCl MAX; Polysciences). After 4 to 6 days of transfection, IgG1, IgG2 and IgG4 antibodies were isolated from cell supernatants using a HiTrap Protein A High Performance column (GE Healthcare), whereas IgG3 antibodies were isolated with a HiTap Protein G High Performance column (GE Healthcare). Antibodies were dialyzed in PBS, overnight at 4 °C, and filter-sterilized through 0.22 µm Spin-X filters. Antibodies were analyzed by size exclusion chromatography (GE Healthcare) and monomeric fractions were isolated in case of aggregation levels > 5%. The concentration of the antibodies was determined by measurement of the absorbance at 280 nm and antibodies were stored at -20 °C until use. The anti-Hla (MEDI4893; patent O/2017/075188A2), which served as a control, was a gift from Dr Alexey Ruzin, MedImmune (AstraZeneca).

Cloning, expression and purification of SpA-B constructs

Codon-optimized gBlocks (IDT) for wild-type B domain of SpA (SpA-B), SpA-B lacking Fc-binding properties (SpA-B^{KK}; Q9K and Q10K mutations) and SpA-B lacking Fab-binding properties (SpA-B^{AA}; D36A and D37A mutations) (**Table S1**), containing a C-terminal LPETG sortagging sequence, were cloned into a modified pRSET-C-HIS vector by Gibson assembly. Recombinant proteins, containing a C-terminal LPETGG-AAA-HHHHHH tag, were generated in E. coli BL21(DE3) and were isolated under native purification conditions using a 5 HisTrap High Performance Column (GE Healthcare) with imidazole gradient (10-250 mM; Sigma-Aldrich). SpA-B constructs were dialyzed in 50 mM Tris 300 mM NaCl pH 8.0, overnight at 4 °C, and stored at -20 °C until use.

Antibody cleavage

Anti-CD52, anti-WTA and anti-Hla IgG1 Fc and $F(ab')_2$ molecules were generated by overnight digestion at 37 °C using 0.25 U/µg FabRICATOR IdeS (Genovis AB). FITC-conjugated rabbit $F(ab')_2$ anti-human C1q was obtained by digestion of FITC-conjugated

rabbit anti-human C1q (Dako) using 1 U/µg of recombinant His-tagged IdeS protease. After incubation of antibody with IdeS for 2 h at 37 °C in the dark, $F(ab')_2$ fragment was purified though HiTrap Protein A High Performance column (GE Healthcare) and HisTrap High Performance column (GE Healthcare).

Antibody labeling

IgG labeling was performed by incubation of antibodies with Alexa Fluor647 NHS Ester (Succinimidyl Ester; ThermoFisher Scientific). In detail, 1 µL of 10 mg/mL Alexa Fluor647 NHS Ester was added to 100 µl of 1 mg/mL antibody in PBS with 0.1 M sodium bicarbonate buffer. After 2 h incubation at RT in the dark, labeled antibodies were separated from the free probe by use of Pierce desalting spin column 0.5 mL (ThermoFisher Scientific), according to the manufacturer's manual. Subsequently, absorbance was measured at 280 nm for protein and 647 nm for the probe, using a Nanodrop.

Binding of antibodies to beads coated with SpA constructs

Dynabeads His-Tag Isolation & Pulldown (Invitrogen) were washed in PBS-TH (PBS, 0.05% (v/v) Tween-20 and 0.5% human serum albumin (HSA)). 50 µL of 30 µg/mL his-tagged SpA-B, SpA-B^{KK} or SpA-B^{AA} was mixed with 1 µL beads for 30 min at 4 °C, shaking (± 700 rpm). After two washes with PBS-TH, 0.05 µL beads were incubated with 2 µg/mL anti-DNP IgG1, anti-DNP IgG3 or anti-WTA IgG1 in 30 µL, for 30 min at 4 °C, shaking (± 700 rpm). A VH3-type antibody, anti-Hla IgG1, was included as a control for Fab binding to both SpA-B and SpA-B^{KK}. Anti-DNP IgG3 was included as a negative control. All antibodies were also incubated with Dynabeads Protein G (Invitrogen) that served as reference binding for all antibodies, including IgG3. Antibody binding was detected using directly labeled IgG or Alexa $Fluor^{647}$-conjugated goat $F(ab')_2$ anti-human kappa (Southern Biotech). Samples were measured using flow cytometry (BD FACSVerse) and data, based on a single bead population, were analyzed using FlowJo software. Mean fluorescence values were expressed relative to the control binding to Protein G beads for each antibody.

Enzyme-Linked Immunosorbent Assay (ELISA)

MaxiSorp plates (Nunc) were coated with 3 ug/mL SpA-B or SpA in 0.1 M sodium carbonate, overnight at 4 °C. The day after, plates were washed three times with PBS-T (PBS, 0.05% (v/v) Tween-20) and blocked with 4% bovine serum albumin (BSA) in PBS-T, for 1 h at 37 °C. The following incubations were performed for 1 h at 37 °C followed by 3 washes with PBS-T. Five-fold serial dilutions (starting from 20 nM) of human monoclonal anti-DNP IgG1, IgG2, IgG3 and IgG4 in 1% BSA in PBS-T were added to the wells. Bound antibodies were detected with horseradish peroxidase (HRP)-conjugated goat $F(ab')_2$ anti-human kappa (Southern Biotech) in 1% BSA in PBS-T and Tetramethylbenzidine (TMB) as substrate. The reaction was stopped with 1N sulfuric acid and absorbance was measured at 450 nm in the iMark™ Microplate Absorbance Reader (Bio-Rad).

DNP-labeled liposomes

DNP-labeled liposomes consisting of 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (DPPE) and 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-[6-[(2,4-initrophenyl)amino]hexanoyl] (DNP-cap-DPPE) were used to generate supported lipid bilayers (SLBs) on mica and SiO_2 substrates.

The lipids were purchased from Avanti Polar Lipids, mixed at a ratio of DPPC:DPPE:DNP-cap-DPPE = 90:5:5 (molar ratio), and dissolved in a 2:1 mixture of chloroform and methanol. After the solvents were rotary-evaporated for 30 min, the lipids were again dissolved in chloroform, which was then rotary-evaporated for 30 min. Drying was completed at a high vacuum pump for 2 h. The lipids were dissolved in 500 µL Milli-Q H_2O while immersed in a water bath at 60 °C, flooded with argon, and sonicated for 3 min at 60 °C to create small unilamellar vesicles. These were diluted to 2 mg/mL in buffer #1 (10 mM HEPES, 150 mM NaCl, 2 mM $CaCl_2$, pH 7.4) and frozen for storage using liquid N_2.

IgG oligomer statistics on DNP-SLBs

For experiments studying the effect of SpA constructs, the incubation step with SpA proceeded for at least 2 h at 37 °C. IgG-RGY was used at a total IgG concentration of 2 µM in the presence or absence of 20 µM SpA-B or SpA. IgG oligomer distributions were then obtained by incubating DNP-SLBs with 33.3 nM of the respective IgG variant for 5 min, and analyzed in a two-step process: individual particle dimensions were determined by HS-AFM, and the respective oligomeric states were further confirmed by observing their decay pattern determined in subsequent forced dissociation experiments as described[7]. In brief, molecules are scanned in a nondisrupting manner to gauge their number, height, and shape. Subsequently, the scanning force exerted by the HS-AFM cantilever tip is increased to dissociate oligomers into their constituent IgGs. Geometric parameters and oligomer decay patterns are combined to assign each IgG assembly its oligomeric state.

DNP-coated beads assays

Dynabeads M-270 Streptavidin beads (Invitrogen) were washed in PBS-TH and incubated, 100 times diluted, with 1 µg/mL of biotinylated 2,4-dinitrophenol (DNP-PEG2-GSGSGSGK(Biotin)-NH_2, Pepscan Therapeutics B.V.) in PBS-TH for 30 min at 4 °C, shaking (± 700 rpm). For each condition, 0.5 µL of beads were used (~3 x 10^5 beads/condition). After two washes with PBS-TH, DNP-coated beads were incubated with 20 nM anti-DNP IgG or 2-fold serial dilutions of anti-DNP IgG (starting from 20 nM IgG), for 30 min at 4 °C, shaking (± 700 rpm). The following incubations steps were performed under shaking conditions (± 700 rpm) for 30 min at 4 °C unless stated otherwise. Additionally, after each incubation, beads were washed twice with PBS-TH or VBS-TH (Veronal Buffered Saline pH 7.4, 0.5 mM $CaCl_2$, 0.25 mM $MgCl_2$, 0.05% (v/v) Tween-20, 0.5% HSA), dependent on buffer used in the following incubation step. For antibody binding detection, IgG-bound DNP-beads were next incubated with 1 µg/mL Alexa Fluor^{647}-conjugated goat $F(ab')_2$ anti-human kappa (Southern Biotech) in PBS-TH. For SpA binding detection, IgG-bound DNP-coated beads were first incubated with 200 nM of recombinant His-tagged SpA-B or SpA (ProSpec) in PBS-TH. Subsequently, beads were incubated with 1 µg/mL chicken anti-HexaHistidine antibody (Nordic) in PBS-TH, followed by incubation with R-Phycoerythrin (PE)-conjugated donkey $F(ab')_2$ anti-chicken (Jackson ImmunoResearch) diluted 1/500 in PBS-TH. For C1q binding experiments, IgG-bound DNP-coated beads were incubated with 1.3 nM of purified C1 (Complement Technology) or C1q Complement Technology) alone, in combination with, or followed by incubation of 200 nM or 4-fold dilutions (starting from 1 µM) of recombinant SpA-B or SpA in VBS-TH, for 30 min at 37 °C. C1q was detected by use of 4 µg/mL FITC-conjugated rabbit $F(ab')_2$ anti-human C1q. The binding of antibody,

SpA and C1q to the beads were detected using flow cytometry (BD FACSVerse) and data were analyzed based on single bead population using FlowJo software.

Depletion of IgG and IgM from human serum

Human serum pooled from 20 healthy donors was depleted of IgG and IgM as previously reported[8]. Briefly, IgG and IgM were captured by affinity chromatography using HiTrap Protein G High Performance column (GE Healthcare) and CaptureSelect IgM Affinity Matrix (ThermoFisher Scientific), respectively. After depletion, antibodies and complement levels were quantified by ELISA and complement activity was measured by classical pathway and alternative pathway hemolytic assays. As IgG and IgM depletion results in partial co-depletion of C1q, ΔIgG/IgM serum was reconstituted with purified C1q (Complement Technology) to physiological levels (70 μg/mL).

Statistical analysis

Statistical analysis was performed with GraphPad Prism v.8.3 software, using unpaired two-tailed t-test or one-way ANOVA as indicated in the figure legends. At least three experimental replicates were performed to allow statistical analysis.

Supplementary Figures

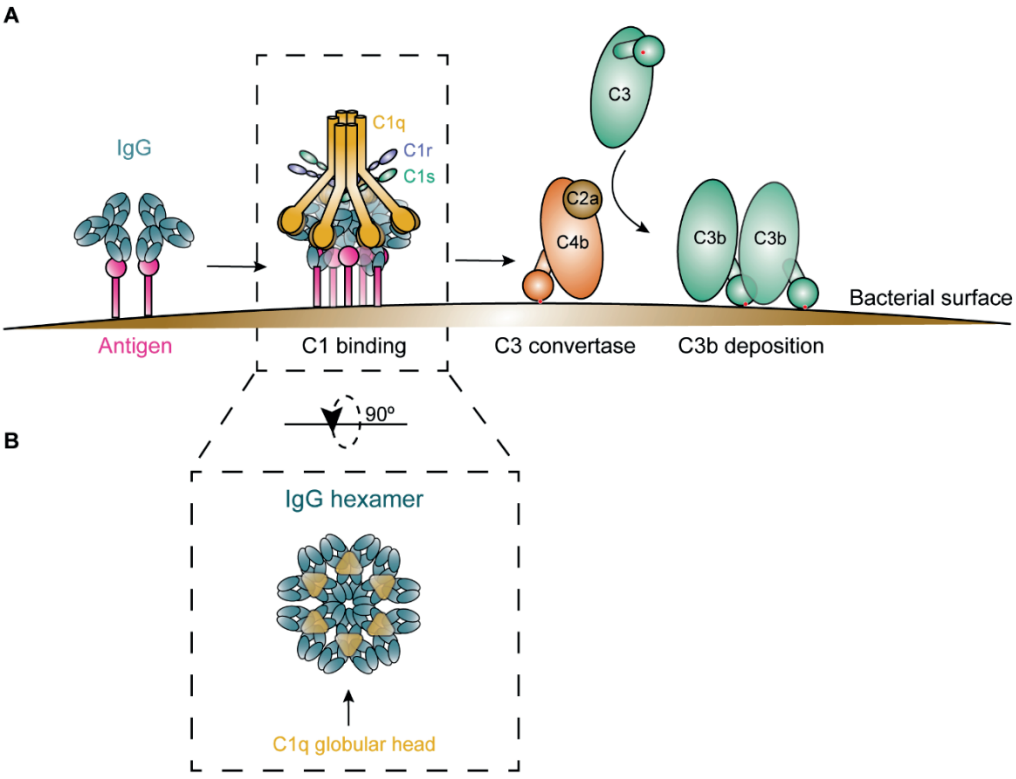


Figure S1 | Graphical representation of complement classical pathway activation. (A) Overview of complement classical pathway initiation. The fully assembled C1 complex binds to immune complexes on the bacterial surface, subsequently, its attached C1r/C1s proteases are activated and cleave C4 and C2 to generate C4b2a (C3 convertase). The C3 convertase cleaves C3 into C3b, which displays a previously hidden thioester that covalently binds to the bacterial surface. **(B)** The IgG hexameric rings, or IgG hexamers, are held together by non-covalent Fc-Fc interactions and form the docking platform for complement C1.

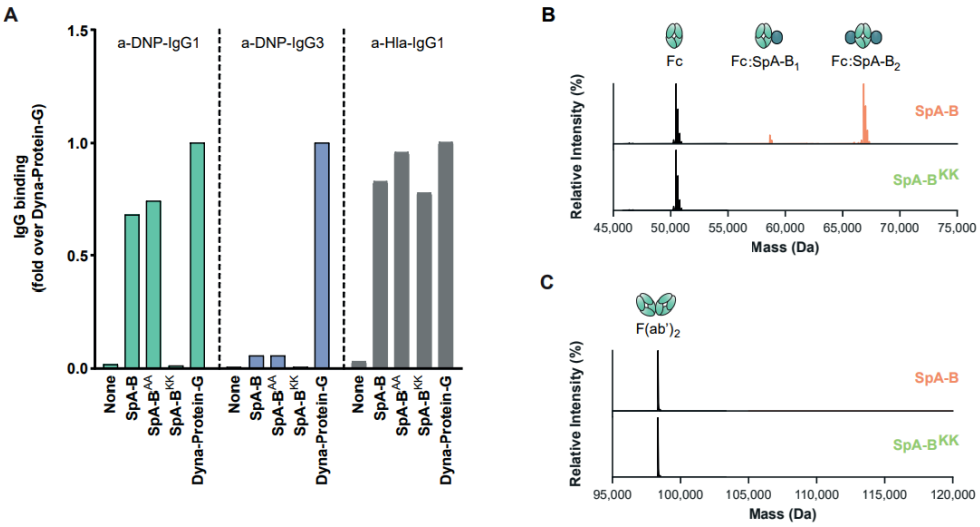


Figure S2 | Human anti-DNP antibodies only bind to SpA via their Fc region. (A) Binding of anti-DNP IgG1 (green), anti-DNP IgG3 (blue) and anti-Hla IgG1 (grey) to beads coated with His-tagged recombinant SpA-B (wild-type), SpA-B^{AA} (that only binds to IgG-Fc region), SpA-B^{KK} (that only binds to IgG-Fab region) or to Protein-G beads. Antibody binding was detected by the use of directly labeled IgG or Alexa Fluor^{647}-conjugated goat F(ab')$_2$ anti-human kappa and was measured by flow cytometry. The anti-Hla IgG1 (VH3 family antibody) served as a control for Fab binding to SpA-B^{KK}. Protein G beads served as universal IgG1 and IgG3 binding control. (B, C) Deconvoluted native mass spectra of the Fc **(B)** and F(ab')$_2$ **(C)** molecules of anti-CD52 IgG1, resulting from IdeS digestion, in the absence (black) or presence of SpA-B (orange) or SpA-B^{KK} (green). SpA-B binds to the Fc region, but not to Fab region. SpA-B^{KK} does not bind to the Fc or Fab region.

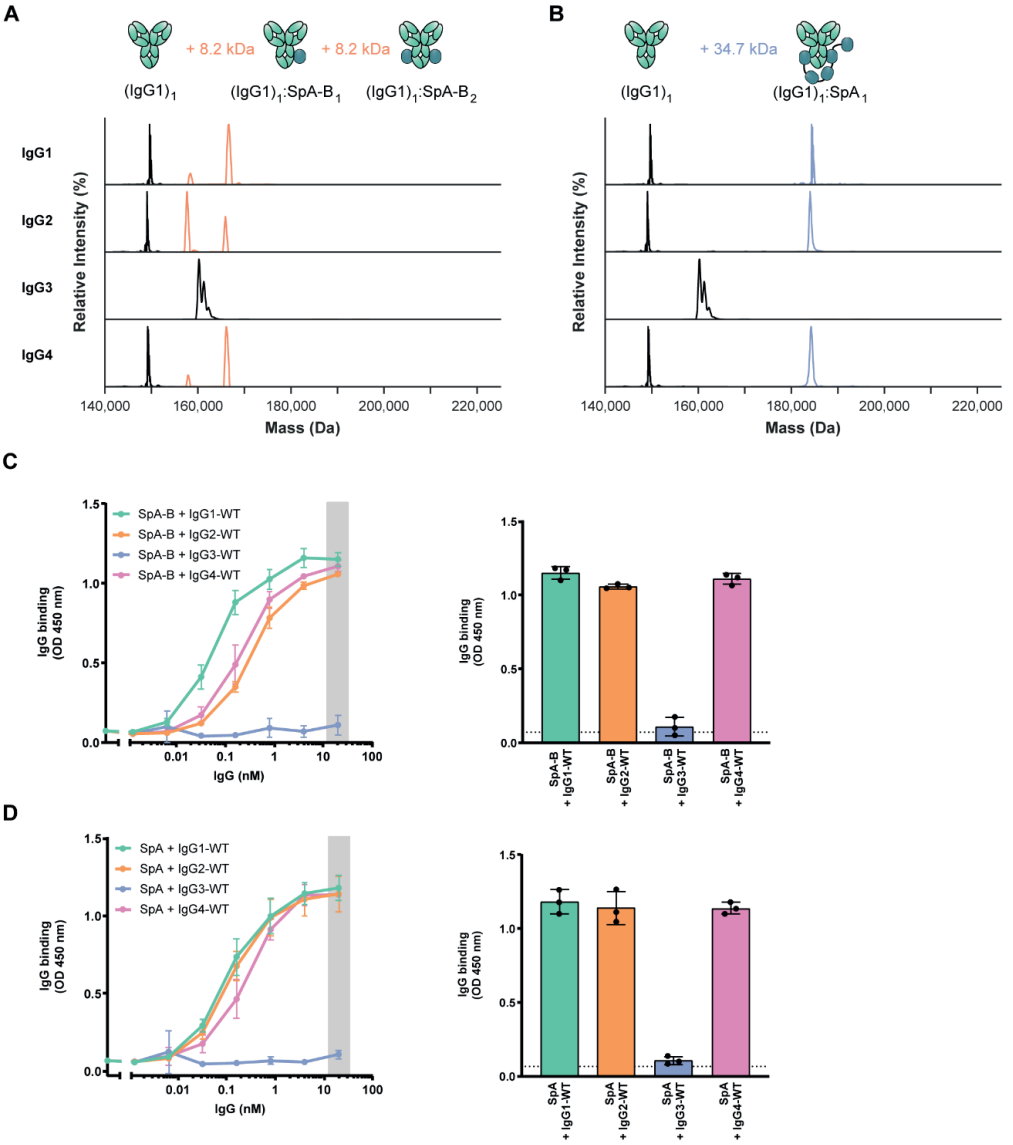


Figure S3 | Human IgG1, IgG2 and IgG4 bind to SpA, but not IgG3. (A, B) Deconvoluted native mass spectra revealing the binding stoichiometry of SpA-B (**A**) and SpA (**B**) with IgG. Incubation with SpA-B (orange) results in mass shifts corresponding to SpA-B binding to IgG1, IgG2 and IgG4 with up to 2:1 stoichiometry whereas SpA (blue) binds principally with 1:1 stoichiometry to the same IgG subclasses. Satellite peaks observed in the mass spectra of IgG3 result from the presence of additional *O*-glycosylation sites on the IgG3 hinge region. The top native MS spectra (human IgG1 binding to SpA-B or SpA) are the same as the spectra of Figure 2. The cartoon regarding $(IgG1)_1$:SpA_1 binding is speculative in the sense it is still unknown which SpA domain contact with which binding site on the IgG molecules. **(C, D)** Binding of the full concentration range of human IgG subclasses to SpA-B (C) or SpA (D) coated on a 96-well plate. IgG binding was detected with HRP-labeled goat $F(ab')_2$ anti-human kappa. Bars represent the same data for the 20 nM IgG concentration only and the black dotted line shows the background absorbance from wells that were not incubated with IgG.

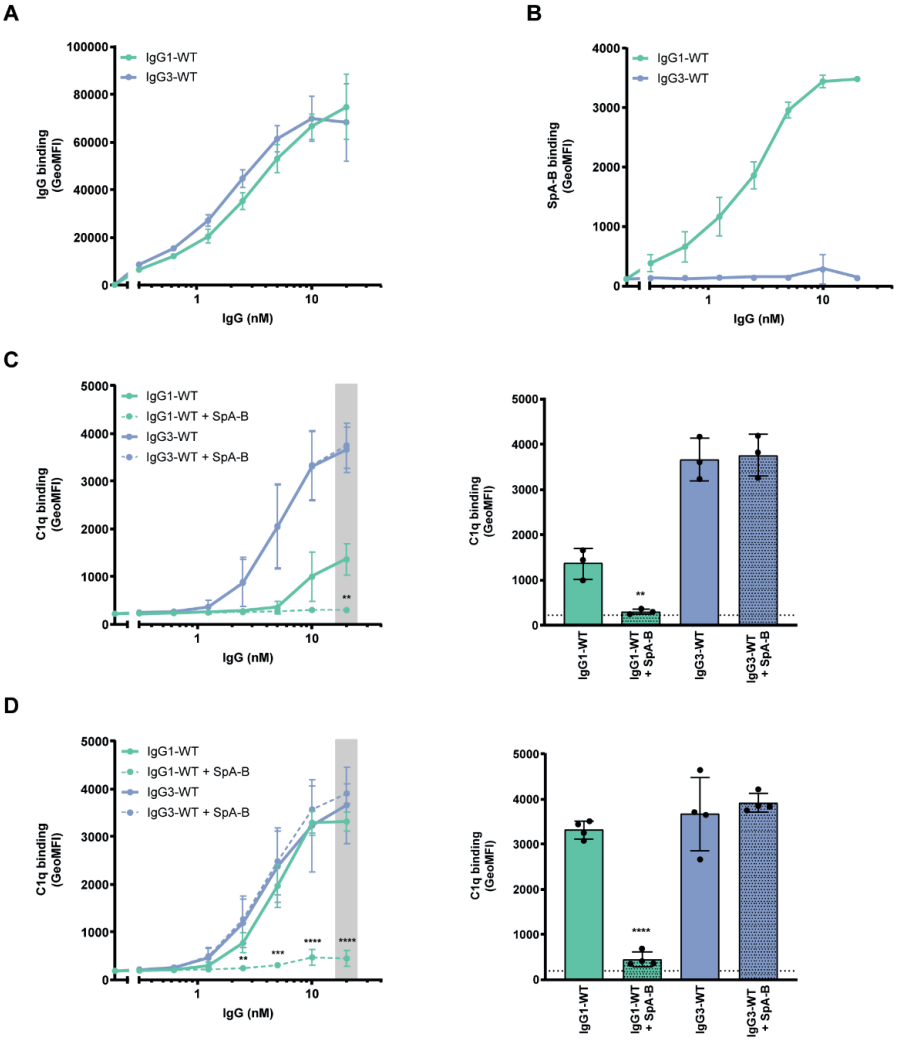


Figure S4 | The binding of SpA-B to antigen-bound IgG decreases C1q binding on target surfaces even if C1q is incubated before SpA-B. (A) Binding of anti-DNP IgG1-WT and IgG3-WT to DNP-coated beads, detected with Alexa Fluor^{647}-conjugated goat F(ab')$_2$ anti-human kappa by flow-cytometry. **(B)** Binding of SpA-B to anti-DNP IgG1-WT and IgG3-WT bound to DNP-coated beads, detected with anti-HexaHistidine chicken antibody by flow cytometry. **(C, D)** C1q binding on IgG1-WT and IgG3-WT bound to DNP-coated beads, detected with FITC-conjugated rabbit F(ab')$_2$ anti-human C1q by flow-cytometry. Buffer (solid lines) or SpA-B (dotted lines) was added to the beads only after C1q (C) or C1 complex (D). Bars represent the same data for the 20 nM IgG concentration only and the black dotted line shows the background fluorescence from beads that were not incubated with IgG. Data are presented as geometric means ± SD of three or four independent experiments. Statistical analysis was performed using an unpaired two-tailed t-test to compare buffer and SpA-B conditions and showed when significant as **P < 0.01, ***P < 0.001 and ****P < 0.0001.

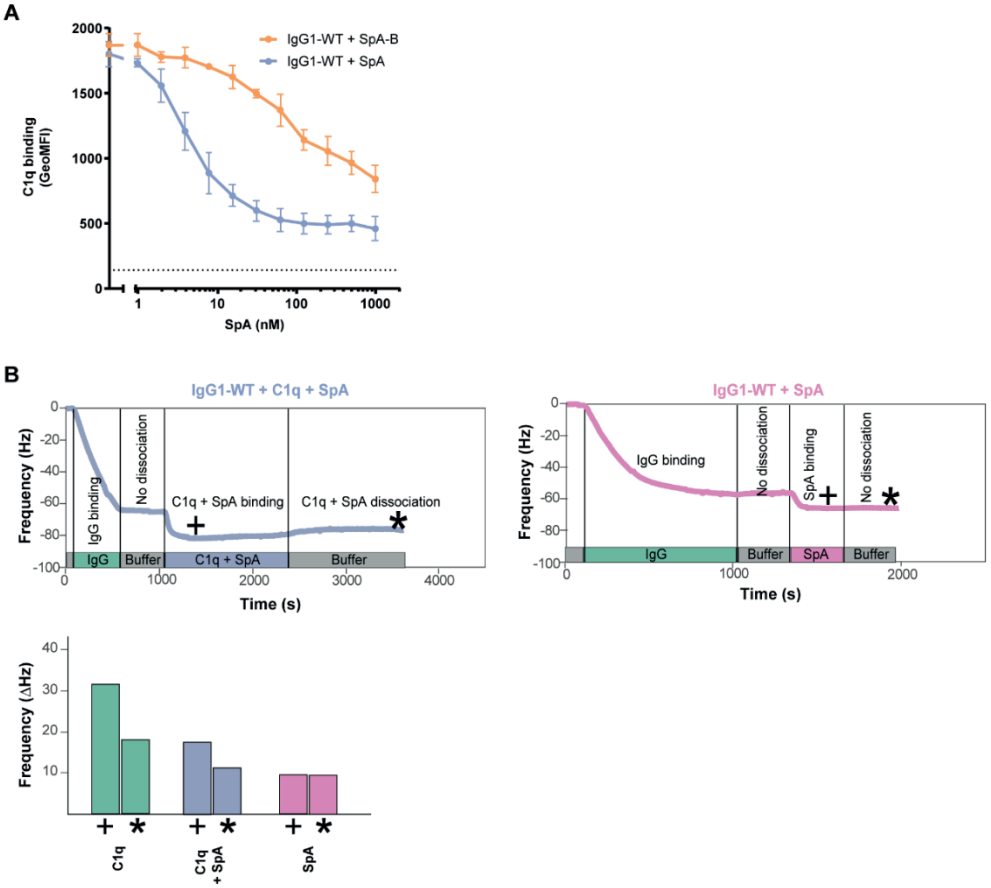


Figure S5 | SpA inhibits binding of C1q to antigen-bound IgGs on target surfaces. (A) C1q binding on anti-DNP IgG1-WT bound to DNP-coated beads after incubation of C1 complex with a concentration range of SpA-B (orange) or SpA (blue), detected with FITC-conjugated rabbit F(ab')$_2$ anti-human C1q antibody by flow cytometry. The black dotted line shows the background fluorescence from beads that were not incubated with IgG. Data are presented as geometric means ± SD of at least three independent experiments. **(B)** QCM sensorgram of C1q binding to anti-DNP IgG1-WT bound to DNP-decorated lipid bilayers in the presence of SpA and SpA binding (in the absence of C1q) to anti-DNP IgG1-WT bound to DNP-decorated lipid bilayers. Bars represent the respective equilibrium level (+) and the level at the end of the dissociation phase (*) (C1q sensorgram is presented in Figure 5A).

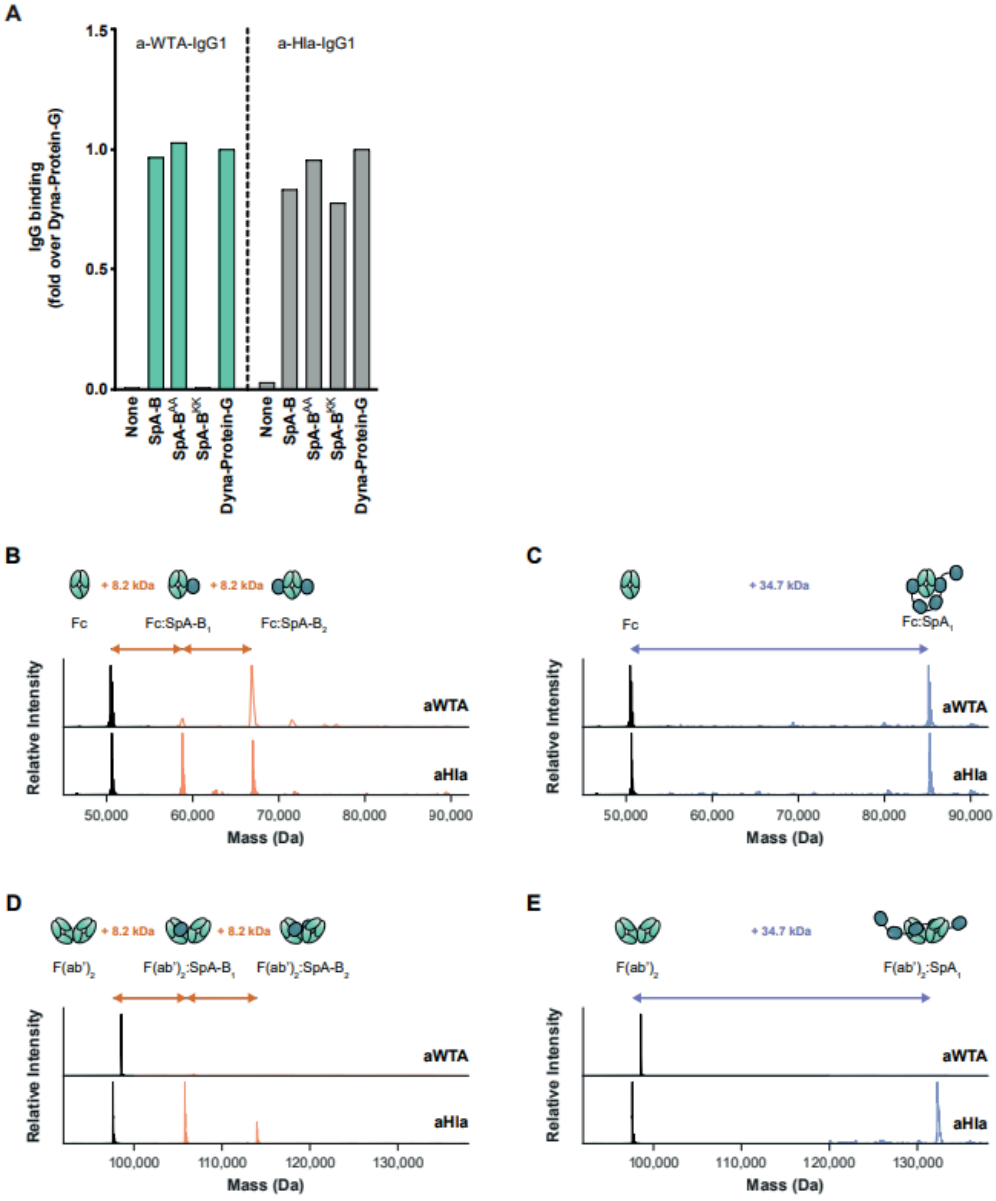


Figure S6 | Human anti-WTA antibodies only bind to SpA via their Fc region. (A) Binding of anti-WTA IgG1 (green) and anti-Hla IgG1 (grey) to beads coated with His-tagged recombinant SpA-B (wild-type), SpA-B^{AA} (that only binds to IgG-Fc region), SpA-B^{KK} (that only binds to IgG-Fab region) or to Protein-G beads. Antibody binding was detected by the use of directly labeled IgG or Alexa $Fluor^{647}$-conjugated goat $F(ab')_2$ anti-human kappa and was measured by flow cytometry. Protein-G beads served as universal IgG1 and IgG3 binding control. The binding data for anti-Hla IgG1 (VH3 family antibody) was also shown in Figure S2. **(B, C)** Deconvoluted native mass spectra of the F(ab')2 molecules of anti-WTA IgG1 or anti-Hla IgG1 resulting from IdeS digestion measured in absence (black) or in presence of SpA-B (orange; B) or SpA (blue; C). Only the F(ab')2 molecule of anti-Hla IgG1 can interact with SpA-B (B) and SpA (C), which bind to these constructs with 2:1 and 1:1 stoichiometry respectively.

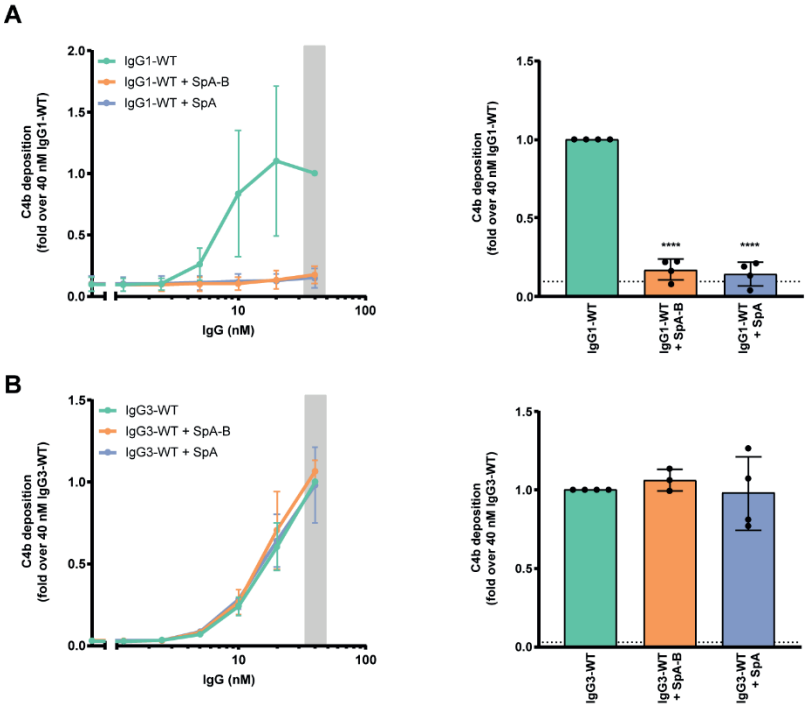


Figure S7 | SpA decreases IgG-mediated C4b deposition on *S. aureus*. (A, B) C4b deposition on Newman*Δspa/sbi* surface after bacteria incubation with IgG1-WT (A) or IgG3-WT (B), 1% ΔIgG/IgM human serum and buffer (green), SpA-B (orange) or SpA (blue), detected with a monoclonal murine anti-human C4d antibody by flow cytometry. Bars represent the same data for the 40 nM IgG concentration only and the black dotted line shows the background fluorescence from bacteria that were not incubated with IgG. Data are presented as fold change over the 40 nM IgG concentration control ± SD of at least three independent experiments. Statistical analysis was performed using one-way ANOVA to compare buffer condition with SpA-B and SpA conditions and showed when significant as ****P < 0.0001.

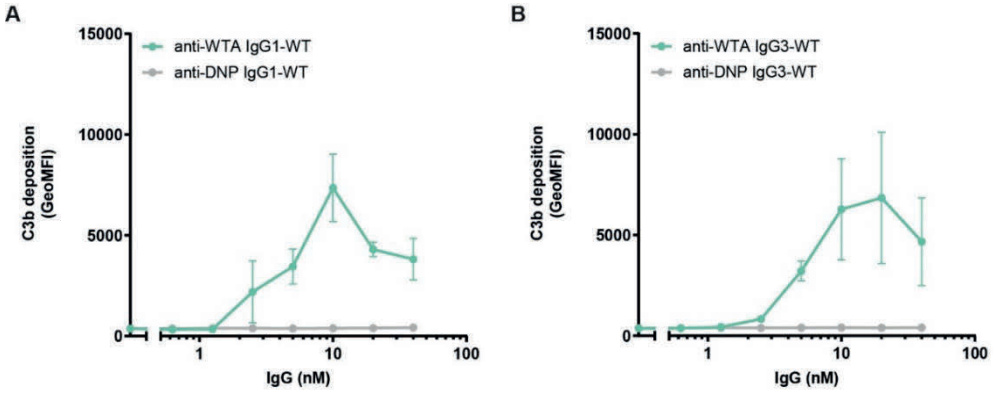


Figure S8 | C3b detection by anti-WTA antibodies reflects the presence of covalently bound C3 products that were deposited upon complement activation. (A, B) C3b deposition on NewmanΔspa/sbi surface after incubation of bacteria with anti-WTA wild-type (anti-WTA; green line) or anti-DNP wild-type (anti-DNP; grey line) antibodies of IgG1 subclass (A) or of IgG3 subclass (B) in presence of 1% ΔIgG/IgM human serum, detected with a monoclonal murine anti-human C3d antibody by flow cytometry. Data are presented as geometric means ± SEM of two independent experiments.

Supplementary Tables

Table S1 | Amino acid sequence of SpA-B constructs. The amino acid substitutions that abrogate the binding of SpA-B^{KK} and SpA-B^{AA} to the Fc and Fab of human IgG, respectively, are underlined and highlighted in bold.

SpA-B construct	Sequence
SpA-B	MADNKFNKEQQNAFYEILHLPNLNEEQRNGFIQSLKDDPSQSANLLAEAKKLNDAQAPK
SpA-B^{KK}	MADNKFNKE**KK**NAFYEILHLPNLNEEQRNGFIQSLKDDPSQSANLLAEAKKLNDAQAPK
SpA-B^{AA}	MADNKFNKEQQNAFYEILHLPNLNEEQRNGFIQSLK**AA**PSQSANLLAEAKKLNDAQAPK

Table S2 | Theoretical and experimental molecular weight values of proteins analyzed by native MS. The mass deviations of ≈ 2.5 kDa for the IgGs are due to the *N*-glycosylations in the Fc region, whereas the 42 kDa mass difference for C1q is expected due to extensive glycosylation of the α, β and γ subunits.

Species	$MW_{theoretical}$ (Da)	$MW_{experimental}$ (Da)	ΔMW (Da)
SpA-B	8,320	8,188	-132*
SpA-BKK	8,320	8,188	-132*
SpA	34,660	34,663	3
IgG1	146,995	149,217	2,222
IgG1-RGY	147,057	149,831	2,774
IgG2	146,428	149,065	2,636
IgG2-RGY	146,490	149,137	2,647
IgG3	157,219	160,113	2,894
IgG3-RGY	157,281	159,929	2,647
IgG4	146,544	149,181	2,637
IgG4-RGY	146,606	149,237	2,631
C1q	421,437	463,591	42,153

* Corresponding to the deletion of Met1 in the protein sequence.

Table S3 | Theoretical and experimental mass values of proteins complexes related to Figure 2A, native mass spectra showing the effect of SpA-B and SpA on IgG1-RGY hexamerization. The theoretical mass of the complex, as calculated using the measured masses of individual subunits, is compared to the experimental mass. For these experiments, instrumental parameters were optimized for the quantification of $(IgG)_6$ complexes.

Species	$M_{theoretical}$ (kDa)	$M_{experimental}$ (kDa)	ΔM (kDa)
$(IgG1)_1$	149.8	149.9	0.07
$(IgG1)_1$:$(SpA\text{-}B)_1$	158.0	158.3	0.32
$(IgG1)_1$:$(SpA\text{-}B)_2$	166.2	166.6	0.39
$(IgG1)_1$:$(SpA)_1$	184.5	184.8	0.31
$(IgG1)_2$	299.7	299.9	0.25
$(IgG1)_3$	449.5	449.9	0.39
$(IgG1)_4$	599.3	599.8	0.51
$(IgG1)_6$	899.0	900.4	1.43

Table S4 | Theoretical and experimental mass values of proteins complexes related to Figure 3A, native mass spectra showing the effect of SpA-B and SpA on $(IgG1\text{-}RGY)_6$:C1q assembly. The theoretical mass of the complex, as calculated using the measured masses of individual subunits, is compared to the experimental mass. For these experiments, instrumental parameters were optimized for the quantification of $(IgG)_6$:C1q complexes.

Species	$M_{theoretical}$ (kDa)	$M_{experimental}$ (kDa)	ΔM (kDa)
$(IgG1)_1$	149.8	150.0	0.13
$(IgG1)_1$:$(SpA\text{-}B)_1$	158.0	158.2	0.22
$(IgG1)_1$:$(SpA\text{-}B)_2$	166.2	166.4	0.22
$(IgG1)_1$:$(SpA)_1$	184.5	184.8	0.35
$(IgG1)_2$	299.7	299.9	0.21
$(IgG1)_3$	449.5	449.7	0.24
C1q	463.6	463.8	0.17
$(IgG1)_6$	899.0	899.9	0.95
$(IgG1)_6$:$(C1q)_1$	1,362.6	1,364.5	1.96

Table S5 | Amino acid sequence of the variable and constant heavy and light chains of antibodies produced in this study[5,6].

	Antibody	Sequence
Variable heavy chain	Anti-DNP IgG	DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYD GSNNYNPSLKNRISITRDTSKNQFFLKLNSVTTEDTATYYCARATYYGNYRGFAY WGQGTLVTVSA
	Anti-WTA IgG	EVQLVESGGGLVQPGGSLRLSCSASGFSFNSFWMHWVRQVPGKGLVWISFTNNEG TTTAYADSVRGRFIISRDNAKNTLYLEMNNLRGEDTAVYYCARGDGGLDDWGQGT LVTVSS
Variable light chain	Anti-DNP IgG	DIRMTQTTSSLSASLGDRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLH SGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPWTFGGGTKLEIK
	Anti-WTA IgG	DIQLTQSPDSLAVSLGERATINCKSSQSIFRTSRNKNLLNWYQQRPGQPPRLLIH WASTRKSGVPDRFSGSGFGTDFTLTITSLQAEDVAIYYCQQYFSPPYTFGQGTKL EIK
Constant heavy chain	IgG1	ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPA VLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCP PCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGV EVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTIS KAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYK TTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
	IgG2	ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPA VLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVERKCCVECPPCP APPVAGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHN AKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKG QPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPP MLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
	IgG3	ASTKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPA VLQSSGLYSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELKTPLGDTTHT CPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPAPELL GGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNAKTKP REEQYNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREP QVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESSGQPENNYNTTPPMLDSD GSFFLYSKLTVDKSRWQQGNIFSCSVMHEALHNRFTQKSLSLSPGK
	IgG4	ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPA VLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYGPPCPSCP APEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVH NAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAK GQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTP PVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSPGK
Constant light chain (*kappa*)	IgG1, 2, 3, 4	RTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQES VTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Supplemental References

1. Diebolder, C.A. et al. Complement is activated by IgG hexamers assembled at the cell surface. *Science* **343**, 1260-1263 (2014).
2. White, K.D., Frank, M.B., Foundling, S. & Waxman, F.J. Effect of immunoglobulin variable region structure on C3b and C4b deposition. *Mol. Immunol.* **33**, 759-768 (1996).
3. Ugurlar, D. et al. Structures of C1-IgG1 provide insights into how danger pattern recognition activates complement. *Science* **359**, 794-797 (2018).
4. Fang, X.T., Sehlin, D., Lannfelt, L., Syvänen, S. & Hultqvist, G. Efficient and inexpensive transient expression of multispecific multivalent antibodies in Expi293 cells. *Biol. Proced. Online* **19**, 11 (2017).
5. Gonzalez, M.L., Frank, M.B., Ramsland, P.A., Hanas, J.S. & Waxman, F.J. Structural analysis of IgG2A monoclonal antibodies in relation to complement deposition and renal immune complex deposition. *Mol. Immunol.* **40**, 307-317 (2003).
6. Lehar, S.M. et al. Novel antibody–antibiotic conjugate eliminates intracellular S. aureus. *Nature* **527**, 323-328 (2015).
7. Strasser, J. et al. Unraveling the Macromolecular Pathways of IgG Oligomerization and Complement Activation on Antigenic Surfaces. *Nano Lett.* **19**, 4787-4796 (2019).
8. Zwarthoff, S.A., Magnoni, S., Aerts, P.C., van Kessel, K.P.M. & Rooijakkers, S.H.M. in The Complement System: Innovative Diagnostic and Research Protocols. (ed. L.T. Roumenina) 21-32 (Springer US, New York, NY; 2021).

CHAPTER 5

CD5L IS A CANONICAL COMPONENT OF CIRCULATORY IGM

Nienke Oskam[1,*], Maurits A. den Boer[2,3,*], Marie V. Lukassen[2,3], Pleuni Ooijevaar-de Heer[1], Tim S. Veth[2,3], Gerard van Mierlo[1], Szu-Hsueh Lai[2,3], Ninotska I.L. Derksen[1], Victor C. Yin[2,3], Marij Streutker[1], Vojtech Franc[2,3], Marta Siborova[2,3], Mirjam Damen[2,3], Dorien Kos[1], Arjan Barendregt[2,3], Albert Bondt[2,3], Johannes B. van Goudoever[4], Carla J.C. de Haas[5], Piet C. Aerts[5], Remy M. Muts[5], Suzan H.M. Rooijakkers[5], Gestur Vidarsson[6,2,3], Theo Rispens[1,#], and Albert J.R. Heck[2,3,#]

[1] Sanquin Research and Landsteiner Laboratory, Department of Immunopathology, Amsterdam UMC, Amsterdam, The Netherlands.

[2] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

[4] Amsterdam UMC, Vrije Universiteit, University of Amsterdam, Emma Children's Hospital, 1105 AZ Amsterdam, The Netherlands

[5] Department of Medical Microbiology, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

[6] Sanquin Research and Landsteiner Laboratory, Department of Experimental Immunohematology, Amsterdam UMC, Amsterdam, The Netherlands.

[*] These authors contributed equally

[#] These authors shared supervision of this study

Published as:

CD5L Is a Canonical Component of Circulatory IgM
Nienke Oskam, Maurits A. den Boer, Marie V. Lukassen, Pleuni Ooijevaar-de Heer, Tim S. Veth, Gerard van Mierlo, Szu-Hsueh Lai, Ninotska I.L. Derksen, Victor C. Yin, Marij Streutker, Vojtech Franc, Marta Siborova, Mirjam Damen, Dorien Kos, Arjan Barendregt, Albert Bondt, Johannes B. van Goudoever, Carla J.C. de Haas, Piet C. Aerts, Remy M. Muts, Suzan H.M. Rooijakkers, Gestur Vidarsson, Theo Rispens, and Albert J.R. Heck
bioRxiv 2023.05.27.542462
DOI: 10.1101/2023.05.27.542462

Abstract

Immunoglobulin M (IgM) is an evolutionary conserved key component of humoral immunity, and the first antibody isotype to emerge during an immune response. IgM is a large (1 MDa), multimeric protein, for which both hexameric and pentameric structures have been described, the latter additionally containing a joining (J) chain. Using a combination of single-particle mass spectrometry and mass photometry, proteomics, and immunochemical assays, we here demonstrate that circulatory (serum) IgM exclusively exists as a complex of J-chain-containing pentamers covalently bound to the small CD5 antigen-like (CD5L, 36 kDa) protein. In sharp contrast, secretory IgM in saliva and milk is principally devoid of CD5L. Unlike IgM itself, CD5L is not produced by B cells, implying that it associates with IgM in the extracellular space. We demonstrate that CD5L integration has functional implications, *i.e.*, it diminishes IgM binding to two of its receptors, the FcαμR and the polymeric Immunoglobulin receptor (pIgR). On the other hand, binding to FcμR as well as complement activation via C1q seem unaffected by CD5L integration. Taken together, we redefine the composition of circulatory IgM as a J-chain containing pentamer, always in complex with CD5L.

Introduction

Immunoglobulin M (IgM) is the first antibody isotype to emerge in ontology and during an immune response. IgM is integral to the initiation of the humoral immune response, but also for maintaining immune homeostasis through the induction of tolerance and for instance the clearance of apoptotic cells[1,2]. It is furthermore a potent activator of the classical pathway of the complement system and regulates antigen presentation and B-cell maturation through interactions with the IgM-specific receptors FcμR[3] and FcαμR[4]. Additionally, secretory IgM plays an important role in mucosal immunity, as the integration of a small joining (J) chain allows it to be transported to mucosal surfaces via the polymeric immunoglobulin receptor (pIgR)[5]. IgM is present throughout our body, with concentrations in serum, human milk and saliva of ca. 1500 mg/L[6], 2.8 mg/L[7] and 1.2 mg/L[8], respectively.

IgM is expressed by B cells as a precursory monomeric (H_2L_2), membrane-bound B cell receptor, which recognizes antigen and relays survival and proliferation signals for maturing B cells. When a B cell switches to IgM secretion through alternative splicing of the μ heavy chains (HC), it can concomitantly co-express the J-chain. Five IgM protomers, each consisting of two heavy chains coupled to two light chains, can combine with a J-chain to form one pentameric molecule, which has for decades been assumed to be the principal arrangement of IgM in circulation[1,9]. Furthermore, it has been shown that if IgM is expressed without J-chain it can assemble into hexamers[10-12]. However, (monoclonal) hexameric IgM observed in circulation seems invariably linked to pathologies such as Waldenström macroglobulinemia or cold agglutinin disease[10,11]. Despite persistent speculation in literature, it is currently unclear what fraction, if any, of normal human circulatory IgM is hexameric[12-14].

Recently, several cryo-EM studies have shed more light on the detailed molecular structure of pentameric IgM. Rather than the previously assumed symmetrical pentameric arrangement, recombinant IgM with J-chain forms an asymmetrical pentamer resembling a hexameric structure, wherein the J-chain bridges the gap in place of a sixth IgM subunit[15-19]. The core of these structures is comprised of an amyloid-like assembly of the J-chain with the C-terminal IgM tailpieces, which are responsible for this isotype's propensity to oligomerize. On top of this assembly, the secretory component (SC) of pIgR can bind IgM through interaction with the J-chain and IgM heavy chains[16,19].

Intriguingly, by using *in vitro* reconstitution, using negative stain EM, it was reported that a protein called CD5-Like molecule (CD5L) can also associate into the gap of murine J-chain-linked pentameric IgM and form a covalent attachment through a disulfide bond[18]. CD5L is a member of the scavenger receptor cysteine-rich (SRCR) family and consists of three SRCR domains in both murine and human CD5L[20]. It was originally discovered as Spα or apoptosis inhibitor of macrophages (AIM), being expressed predominantly by macrophages[21]. In circulation, it has been speculated that CD5L binds to IgM to avoid renal excretion[22], whereby the protein is released under certain conditions. However, it is unclear when, where and how frequently CD5L incorporation happens. Disconcertingly, reported circulatory CD5L concentrations differ wildly, ranging from 0.1 to 60 mg/L (*i.e.*, ca. 500-fold)[20,22-24], possibly reflecting limitations to detect either the free or putatively

IgM-bound form. While free CD5L has been reported to induce a plethora of immunomodulatory effects, the exact functions of this protein remain largely unknown[25-27]. Above all, the incidence and significance of the IgM-bound form of CD5L are currently undetermined. Here, we characterize human circulatory IgM for its structural composition and association with CD5L. Combining molecular biology and mass spectrometry (MS)-based techniques, we redefine the circulatory IgM complex and explore the role of CD5L in complement activation and Fc receptor binding.

Results

CD5L occurs synchronously with IgM in circulation

We started our characterization of IgM and CD5L by tracking the abundances and associations of these proteins in serum from healthy donors (n = 42). To investigate the abundance of the individual protein chains (Igμ, J, CD5L) in a way that is unbiased to structure and complex formation, we used bottom-up proteomics with label-free quantitation (**Figure 1A and S1**). Using this approach, we detected an average total CD5L serum concentration of 1.7 μM or 60 mg/L, consistent with the higher end of literature values[20]. Remarkably, we found that CD5L levels correlated very tightly with those of the IgM heavy chain constant region (IgμC, r = 0.98), hinting at high levels of circulating IgM-CD5L complexes. Furthermore, the observed CD5L/IgμC molecular ratio was consistently ca. 0.15, translating to close to one CD5L molecule per IgM pentamer.

To further investigate a putative structural relationship between IgM and CD5L, we next subjected pooled sera to size-exclusion chromatography (SEC) MS. In this approach, bottom-up proteomics was applied to SEC fractions to generate chromatograms of individual proteins, revealing complexation based on synchronous elution. Strikingly, the chromatogram of CD5L revealed essentially complete (> 99%) synchronous elution with IgM as well as with the J-chain in the higher MW fractions around 1 MDa (**Figure 1B**). No secondary CD5L elution peak was observed in lower MW fractions, indicating that the bulk of CD5L is in association with IgM and the J-chain. Also in SEC MS, we observed an intensity ratio of about 1:1:10 between CD5L, J-chain and IgμC, again hinting towards a 1:1:1 molecular ratio of pentameric IgM, J-chain and CD5L in circulation.

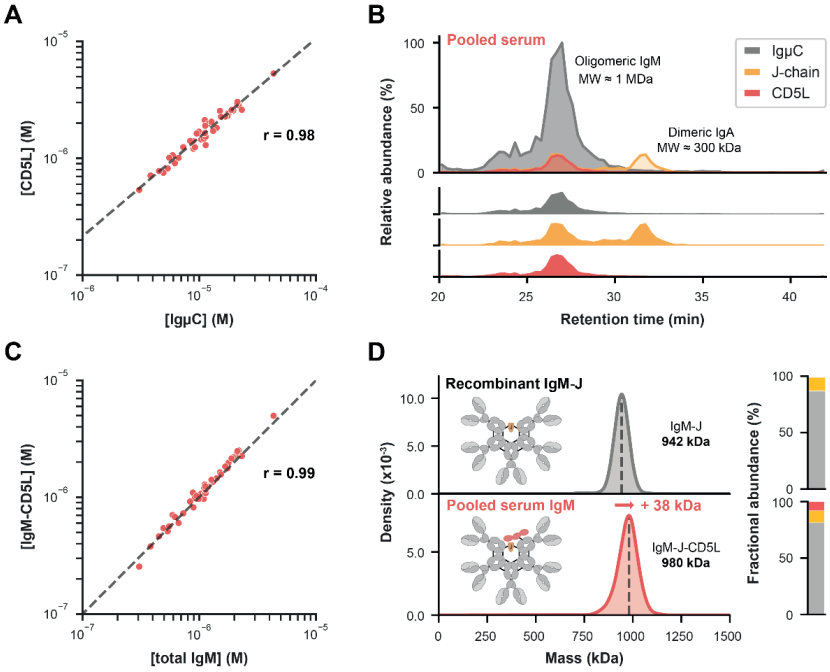


Figure 1 | All circulatory IgM is pentameric with J-chain and CD5L incorporated. (A) Total levels of CD5L and IgM heavy chain (IgμC) show a strong correlation in serum (r = 0.98). Absolute concentrations were assessed by label-free quantitation by proteomics. The molecular ratio between CD5L and IgμC was approximately 1:7, advocating the incorporation of one CD5L molecule per IgM pentamer. The dotted line indicates a linear regression model fitted to logarithmically scaled data. **(B)** In complex-centric protein profiling of pooled serum, CD5L elutes exclusively (> 99%) with IgμC and J-chain as a single complex of ca. 1 MDa. Shown are the elution profiles of the individual protein chains obtained from proteomics on individual size-exclusion chromatography (SEC) fractions. The secondary elution peak of the J-chain corresponds to incorporation in dimeric IgA. **(C)** Levels of IgM-CD5L complexes and total IgM ELISA similarly show a high correlation in serum (r = 0.99). The ELISA makes use of a mAb (5B5) that can recognize CD5L when bound to IgM. Quantification was based on a recombinantly produced IgM-J-CD5L complex standard (Figure 2, S5). The ratio between IgM-CD5L complexes and total IgM was approximately 1:1, implying that all serum IgM incorporates a CD5L molecule. The grey line indicates this 1:1 molecular ratio. **(D)** The principal configuration of circulatory IgM is a pentamer with J-chain and CD5L. Combining mass measurement by CDMS (left) with proteomics (right) confirmed that recombinant IgM-J (targeting wall teichoic acid of *S. aureus*) is a pentamer with J-chain ($(IgM)_5$:$(J)_1$) (top). IgM purified from pooled serum was similarly homogeneous (bottom), though the average mass was shifted by + 38 kDa and CD5L was detected by proteomics. This matches the mass increase expected for the incorporation of one CD5L molecule to form ($(IgM)_5$:$(J)_1$:$(CD5L)_1$).

IgM in circulation is principally a J-chain-linked pentamer with CD5L

Having demonstrated that CD5L is tightly connected to IgM in circulation, we next proceeded to characterize the exact molecular composition of circulatory IgM. First, to directly measure and quantify levels of CD5L-IgM complexes, we set up an ELISA that specifically measures these complexes. For this, we generated a dedicated mAb (5B5) that recognizes CD5L when it is bound to IgM (**Figure S2**). Using this approach, we found that in the serum of 42 healthy donors, the levels of IgM-bound CD5L and total IgM also showed a near-perfect correlation (r = 0.99) and a molecular ratio of 1 (**Figure 1C**), further implying that IgM exists mainly as a CD5L-containing complex in circulation. In contrast,

employing a set-up that solely detects unbound CD5L, we found those levels averaged 0.7 mg/L, thus making up only a small fraction (~ 1%) of the total circulating CD5L population (**Figure S2D**).

To elucidate the exact molecular composition of circulatory IgM, we next subjected IgM purified from pooled serum and the plasma of individual donors to two single-particle mass measurement techniques, native charge detection (CD) MS[28] and mass photometry[29]. Combining these precise intact molecular weight measurements with bottom-up proteomics to identify constituent proteins, allowed us to establish the exact stoichiometries and possible co-occurrences of different oligomeric states (**Figure 1D** and **S3**). For comparison we also analyzed a recombinant IgM-J and observed it to be exclusively a J-chain-linked pentamer with a measured mass of 942 kDa ($(IgM)_5$:$(J)_1$). For each serum or plasma IgM sample, we consistently observed a mass of about 40 kDa higher, a shift that closely matches the incorporation of one 36 kDa CD5L molecule. Simultaneously, proteomics analyses on the same circulatory IgM samples consistently revealed the presence of CD5L at a molecular ratio of about 1:1:10 (CD5L:J-chain:IgμC). Notably, circulatory IgM was also highly homogeneous in mass and therefore oligomeric state, without evidence for the existence of IgM hexamers in any of the samples analyzed (abundance < 1%) (**Figure 1**). We thus conclude that the canonical form of human circulatory IgM is a J-chain-linked pentamer with one CD5L molecule ($(IgM)_5$:$(J)_1$:$(CD5L)_1$).

CD5L is not expressed by B cells and is covalently integrated into IgM via CD5L-Cys191

The near-complete integration of CD5L into IgM as observed here would be best explained by secretion from B cells as a fully formed complex. However, the expression of CD5L is reported to be mostly restricted to macrophages[21,30]. To investigate the possibility of co-expression of CD5L and IgM in B cells, we cultured naïve and memory B cells, both in a T cell-dependent and T cell-independent manner and analyzed IgM in the supernatants (**Figure 2A**). The secreted IgM was completely devoid of CD5L, demonstrating that B cells do not produce CD5L. Recombinant co-transfection of IgM-J and CD5L within the same cell proved to be challenging and did not lead to efficient complex formation. In contrast, co-culture of IgM-J-expressing and CD5L-expressing cells reliably produced homogenous IgM-J-CD5L (**Figure S4**). This implies that complex formation of B cell-derived IgM with CD5L occurs in the extracellular space after both IgM-J and CD5L have been expressed independently.

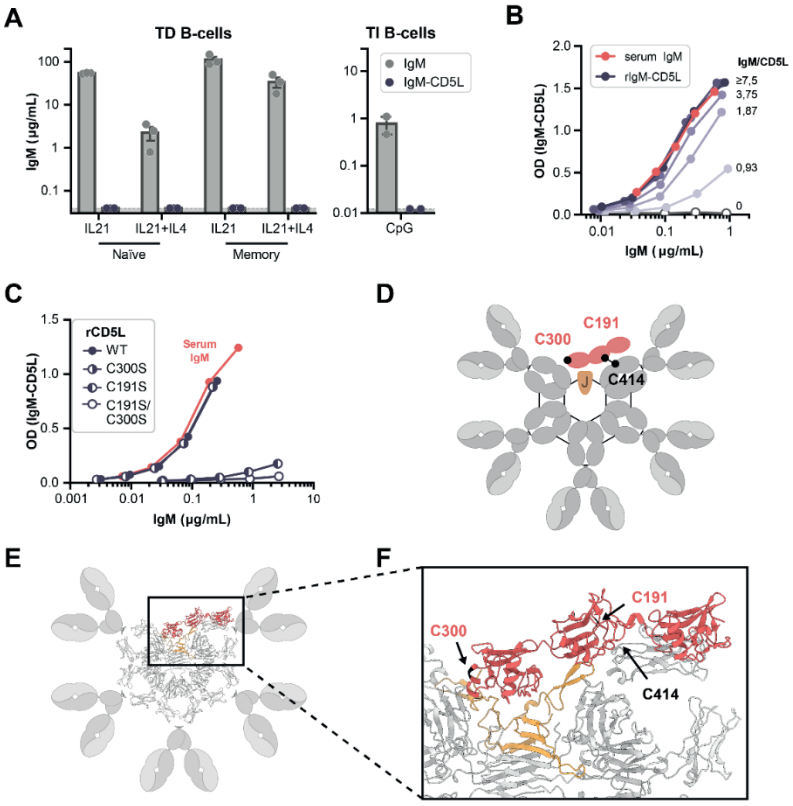


Figure 2 | Formation of IgM-CD5L complexes. (A) Isolated human peripheral blood B cells were cultured *in vitro* with both T-dependent (TD) and independent (TI) stimuli, after which supernatants were analyzed for secreted IgM and IgM-CD5L complexes by ELISA. Only IgM devoid of CD5L was detected, irrespective of culture conditions. Data shown is from 2 and 3 different donors. **(B)** *In vitro* generation of recombinant IgM-CD5L complexes. Different molar ratios of CD5L to IgM (clone 2D5 [14]) were incubated in the presence of 0.1 mM GSH/GSSH. Complex formation was assessed by ELISA. With a molar excess of CD5L, saturation was observed. Representative plots of n = 2 experiments. **(C)** Human CD5L contains two predicted unpaired cysteines, C191 and C300, that could potentially interact with IgM. Recombinant CD5L was produced with C191S or C300S mutations or both and were tested for their ability to form complexes with IgM as determined by ELISA. C191S disrupts complex formation, whereas C300S does not. Representative plots of n = 2 experiments. **(D)** Schematic representation of IgM-CD5L complex with highlighted C191 and C300 of CD5L and C414 of IgM-Fc. E) Proposed structural model of IgM-CD5L complex. AlphaFold2 model of CD5L and J-chain was fitted into the structure of IgM core/Fc region with J-chain (PDB: 8ADY). F) Detailed view on CD5L and J-chain within IgM.

To further clarify the binding mechanism of CD5L to IgM, we recombinantly produced CD5L (rCD5L) and pentameric IgM-J to study CD5L incorporation *in vitro* (**Figure 2B** and **S5**). We observed efficient association of rCD5L with IgM under mild reducing conditions, implicating disulfide bond formation between IgM and CD5L to be an integral part of complex formation. Saturation of IgM binding was reached upon incubation with an excess of CD5L. These binding experiments were repeated with CD5L purified from serum IgM (sIgM-CD5L) and with recombinant IgM-Fc, producing similar results (**Figure S5**).

Integral complex formation was confirmed by shielding of CD5L epitopes for two distinct anti-CD5L mAbs (10D11, 7E4) that do not bind serum-derived IgM (**Figure S2E**).

Based on a structure prediction by Alphafold2[31], human CD5L contains two unpaired cysteines, C191 and C300. To assess their role in IgM-J-CD5L complex formation, we produced recombinant hCD5L variants with these cysteines mutated to serines and tested their ability to associate with IgM-J (**Figure 2C**). Complex formation was severely reduced by a C191S mutation, but essentially unaffected by a C300S mutation. This suggests that only C191 could be disulfide-linked to IgM. This data is in line with the earlier proposed IgM-CD5L linkage via the homologous C194 in experiments wherein murine IgM was reconstituted *in vitro* with murine CD5L[18] (**Figure 2D** and [18]). To further elucidate the interaction, we next subjected CD5L, J-chain and two IgM-Fc regions to structure prediction by Alphafold2 multimer[31]. The model positioned CD5L in the gap of the IgM-Fc pentamer but somewhat sticking out of the IgM-Fc plane (**Figure 2E**, **F**). The SRC3 domain is predicted to be positioned closest to the junction formed by the J-chain, in between the N- and C-terminal loops of the J-chain. The SRC2 domain is located above the cleft between the Fc-Cμ4 and Fc-Cμ3 domains and the N-terminal loop of the J-chain, positioning C191 in close proximity to an unpaired C414 of IgM. The SRC1 domain is expected to be flexible. Taken together, CD5L thus binds the gap of IgM-J and involves disulfide bond formation requiring C191.

IgM-J-CD5L equally activates complement but is less potent in pIgR and FcαμR binding

We next assessed the putative functional consequences of CD5L incorporation into IgM. First, we investigated a potential effect in the activation of the classical pathway of the complement system, for which IgM can be a potent activator (**Figure 3A** and **S6**). We used a recombinant monoclonal IgM antibody specific for biotin with and without CD5L and assessed its ability to induce C3b deposition in ELISA and cell lysis of biotinylated human red blood cells[14]. Both rIgM-J and rIgM-J-CD5L equally induced C3b deposition and cell lysis. We likewise investigated a potential effect of CD5L on C3b deposition on bacteria, using a monoclonal antibody that binds the wall teichoic acid glycopolymer on *S. aureus*[32], but again found no differences in activities (**Figure S6D**).

We next proceeded to characterize the interaction of IgM-J-CD5L with IgM Fc receptor proteins (FcRs) in ELISA. These included pIgR, FcμR and FcαμR, which are involved in mucosal transport[5], regulation of lymphocyte responses[3], and immune complex uptake by antigen-presenting cells[4], respectively. To this end, we measured the binding of recombinant and serum-derived IgM complexes to FcRs (**Figure 3B** and **S7A**). We found that the binding of rIgM-J-CD5L to both pIgR and FcαμR was reduced compared to rIgM-J, while binding to FcμR was unaffected. Also serum-derived IgM (containing CD5L) displayed reduced binding. Binding to pIgR was also assessed using rIgM-Fc and in a reverse binding experiment, both with results corroborating our initial findings (**Figure S7B**). To confirm that these effects were caused by CD5L, we also tested whether release of CD5L from serum IgM enhances binding to pIgR and FcαμR. Selective release of CD5L from serum IgM upon limited reducing conditions proved feasible (**Figure 3C** and **S7C**).

This resulted in increased binding to the pIgR and FcαμR, while binding to the FcμR was not affected (**Figure S7D**). Combined, CD5L incorporation reduces the binding of IgM-J to pIgR and FcαμR, but has no substantial effects on binding to FcμR or complement activation by IgM-J.

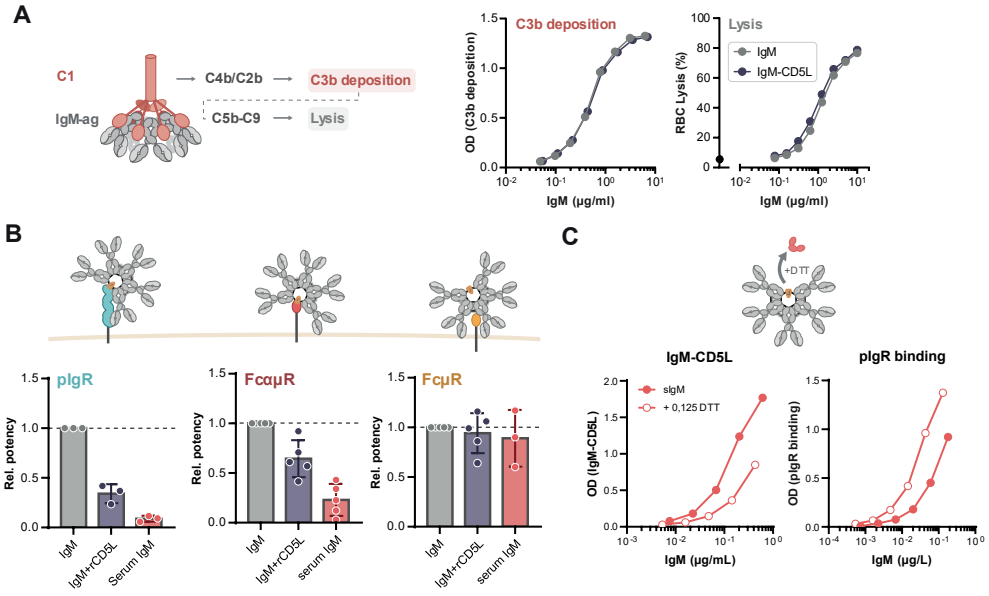


Figure 3 | CD5L in IgM decreases pIgR and FcαμR binding; complement activation and FcμR binding unaltered. (A) Complement activation by IgM was assessed by measuring C3b deposition or red blood cell (RBC) lysis upon binding of recombinant anti-biotin IgM +/- CD5L to plates coated with biotinylated albumin, or biotinylated RBCs, respectively. Representative plots of n = 3 experiments. **(B)** Binding of recombinant IgM (2D5) +/- CD5L and serum IgM to different IgM-binding receptors. Whereas binding to FcμR is unaffected by CD5L, binding to both pIgR and FcαμR is reduced for CD5L-containing IgM. Data from n = 3-5 experiments. **(C)** Upon limited reduction of serum IgM, selective release of CD5L proved feasible (left panel) which resulted in increased pIgR binding (right panel). Data representative of n = 3 experiments. Bars and error bars represent mean and S.E., respectively.

Secretory IgM is principally devoid of CD5L

Secretory IgM is known to be associated with the secretory component (SC), the extracellular portion of the pIgR, which is cleaved from the cells enzymatically after transcytosis to mucosal apical sides. Based on the reduced reactivity of IgM-J-CD5L with pIgR, we hypothesized that, unlike circulatory IgM, secretory IgM may be devoid of CD5L. In stark contrast to serum, we found that CD5L levels in saliva and milk determined by proteomics were much lower and did not correlate with IgμC (**Figure S8**). In agreement, ELISA detection of IgM-bound CD5L in secretory fluids was also much lower than in serum and did not match that of total IgM. Instead, we found a high abundance of SC and a 1:1 molecular ratio between IgM-bound SC and total IgM. In contrast to circulatory IgM, secretory IgM is mostly devoid of CD5L and contains the SC instead.

Secretory IgM is principally devoid of CD5L

Secretory IgM is known to be associated with the secretory component (SC), the extracellular portion of the pIgR, which is cleaved from the cells enzymatically after transcytosis to mucosal apical sides. Based on the reduced reactivity of IgM-J-CD5L with pIgR, we hypothesized that, unlike circulatory IgM, secretory IgM may be devoid of CD5L. In stark contrast to serum, we found that CD5L levels in saliva and milk determined by proteomics were much lower and did not correlate with IgμC (**Figure S8**). In agreement, ELISA detection of IgM-bound CD5L in secretory fluids was also much lower than in serum and did not match that of total IgM. Instead, we found a high abundance of SC and a 1:1 molecular ratio between IgM-bound SC and total IgM. In contrast to circulatory IgM, secretory IgM is mostly devoid of CD5L and contains the SC instead.

Discussion

In this study, we redefined the molecular composition of IgM and revealed its intimate link with the scavenger protein CD5L in circulation. Strikingly, we have now shown that the principal form of circulatory IgM is a J-chain-linked pentamer in complex with CD5L. Conversely, while secretory IgM also consists of J-chain linked pentamer, it is mostly devoid of CD5L but associated with the SC. We speculate that this distinction has eluded discovery due to the relatively small size of CD5L, its assembly after secretion of IgM-J, and the lack of dedicated assays.

While CD5L has been known to co-purify with IgM, its relatively small size (about 4% of the IgM mass) makes it easy to overlook its presence and quantify its abundance. Equally, immunoassays are likely to have underestimated CD5L levels since they were not optimized to include the detection of the IgM-bound form. Nonetheless, several recent plasma proteomics studies revealed high concentrations of total CD5L and a tight correlation with IgM abundance[33-35], in line with our data and an earlier study by Arai *et al.*[22]. In contrast to the J-chain, CD5L is not B-cell derived and is thus integrated into IgM after its expression by a different cell type. Many studies on antibodies have traditionally been performed using immortalized hybridoma B cells and may have wrongly inferred that the resulting CD5L-devoid IgM molecules are representative of circulatory IgM.

While several impressive high-resolution cryo-EM structures of IgM have been published in recent years[15-17,19], the authentic picture of circulatory IgM is thus still incomplete. Most of these studies used *in vitro*-produced IgM without CD5L, with only one negative-stain EM study describing the putative binding of murine CD5L to murine IgM, when reconstituted *in vitro*[18]. A very recent cryo-EM study resolved a structure of serum-derived IgM[17], but surprisingly, also here, CD5L is lacking. We speculate that this may be due to the use of a myeloma-derived sample in this study, wherein IgM may be overproduced to a level where CD5L incorporation is much lower. Further structural studies on healthy human IgM that does include CD5L are thus highly encouraged.

Also, the role of (free) CD5L may need some rethinking. Both anti-and proinflammatory effects have been attributed to CD5L[25,27], but often such experiments were done in the

absence of IgM. High concentrations of CD5L may be produced locally within tissues, but once bound to IgM, CD5L likely needs to dissociate in order to exert its functions[36]. Of note, mice lacking IgM or J-chain show markedly reduced levels of CD5L, emphasizing the tight link between the two[22]. We hypothesize that dissociation of CD5L from IgM may occur for instance at inflamed sites, possibly through released glutathione or thioredoxin. Indeed, CD5L levels have been reported to be increased in a number of inflammatory disorders[23,24,37-42]. These dynamics may resemble redox-dependent association/dissociation seen for IgG4 Fab arm exchange[43]. Alternatively, inflammation may induce increased expression of CD5L by macrophages. Still, the findings of these studies need to be rethought based on our finding that CD5L preliminary associates with IgM in circulation.

Several other exciting questions remain to be addressed. In particular, how and where in the body IgM unites with CD5L during homeostasis as well as during an active (humoral) immune response, and how the complex finds its way into circulation or mucosa? Furthermore, it is currently unclear if the clonal repertoire of IgM is skewed with regard to circulatory or secretory IgM, and how that relates to CD5L incorporation. In theory, this might be akin to the distinction seen in circulatory monomeric IgA and secretory dimeric/polymeric IgA which contains J-chain[44].

Regardless, functional separation of IgM with or without CD5L in the circulation or mucosa, respectively, might establish as either a function of the observational decreased affinity to pIgR, and/or less time for IgM produced in local mucosal tissues to acquire CD5L. In addition, these molecules are likely to be functionally distinct due to altered FcαμR binding, which is thought to play a role in immune complex uptake and antigen presentation thereby shaping the adaptive immune response. Therefore, it will be highly important that studies towards the role of IgM in modulating immune responses will explicitly take CD5L into consideration. Also, IgM is more and more considered as an alternative format for therapeutic antibodies, for which incorporation of CD5L might be a relevant consideration. We conclude that these results show that the molecular structure of IgM needs to be redrafted (**Figure 4**), either containing CD5L in circulation or SC on mucosa.

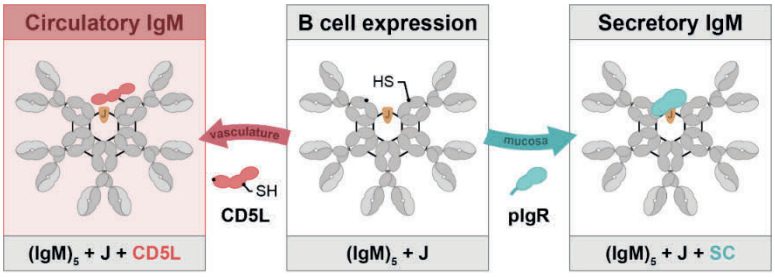


Figure 4 | Graphical summary. J-chain coupled IgM pentamers produced in B cells exclusively engages with CD5L when destined to be secreted into the bloodstream, whereas they attach to the secretory component of pIgR *en route* into secretory biofluids such as milk and saliva.

Funding

We acknowledge support from the Netherlands Organization for Scientific Research (NWO) funding the X-omics Road Map program (Project 184.034.019). Part of this work was supported by the Dutch Arthritis Foundation grant 17-2-404. AJRH acknowledges further support by NWO through the Spinoza Award SPI.2017.028. MVL acknowledges fellowship support from the Independent Research Fund Denmark (Project 9036-00007B).

Acknowledgements

We thank Mads Larsen and Jana Koers for valuable help with the B cell cultures, and René Toes for critical review of the manuscript.

References

1. Blandino, R. & Baumgarth, N. Secreted IgM: New tricks for an old molecule. *Journal of Leukocyte Biology* **106**, 1021-1034 (2019).
2. Grönwall, C. & Silverman, G.J. Natural IgM: Beneficial Autoantibodies for the Control of Inflammatory and Autoimmune Disease. *J. Clin. Immunol.* **34**, 12-21 (2014).
3. Kubagawa, H. et al. The Long Elusive IgM Fc Receptor, FcμR. *J. Clin. Immunol.* **34**, 35-45 (2014).
4. Yoo, E.M., Trinh, K.R., Lim, H., Wims, L.A. & Morrison, S.L. Characterization of IgA and IgM binding and internalization by surface-expressed human Fcα/μ receptor. *Mol. Immunol.* **48**, 1818-1826 (2011).
5. Turula, H. & Wobus, C.E. The Role of the Polymeric Immunoglobulin Receptor and Secretory Immunoglobulins during Mucosal Infection and Immunity. *Viruses* **10**, 237 (2018).
6. Gonzalez-Quintela, A. et al. Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clinical and Experimental Immunology* **151**, 42-50 (2007).
7. Czosnykowska-Łukacka, M., Lis-Kuberka, J., Królak-Olejnik, B. & Orczyk-Pawiłowicz, M. Changes in Human Milk Immunoglobulin Profile During Prolonged Lactation. *Frontiers in Pediatrics* **8** (2020).
8. Michaud, E., Mastrandrea, C., Rochereau, N. & Paul, S. Human Secretory IgM: An Elusive Player in Mucosal Immunity. *Trends Immunol.* **41**, 141-156 (2020).
9. Keyt, B.A., Baliga, R., Sinclair, A.M., Carroll, S.F. & Peterson, M.S. Structure, Function, and Therapeutic Use of IgM Antibodies. *Antibodies* **9**, 53 (2020).
10. Hughey, C.T., Brewer, J.W., Colosia, A.D., Rosse, W.F. & Corley, R.B. Production of IgM Hexamers by Normal and Autoimmune B Cells: Implications for the Physiologic Role of Hexameric IgM1. *J. Immunol.* **161**, 4091-4097 (1998).
11. Eskeland, T. & Christensen, T.B. IgM Molecules with and without J Chain in Serum and after Purification, Studied by Ultra-centrifugation, Electrophoresis, and Electron Microscopy. *Scand. J. Immunol.* **4**, 217-228 (1975).
12. Brewer, J.W., Randall, T.D., Parkhouse, R.M.E. & Corley, R.B. IgM hexamers? *Immunol. Today* **15**, 165-168 (1994).
13. Collins, C., Tsui, F.W.L. & Shulman, M.J. Differential activation of human and guinea pig complement by pentameric and hexameric IgM. *Eur. J. Immunol.* **32**, 1802-1810 (2002).
14. Oskam, N. et al. At Critically Low Antigen Densities, IgM Hexamers Outcompete Both IgM Pentamers and IgG1 for Human Complement Deposition and Complement-Dependent Cytotoxicity. *J. Immunol.* **209**, 16-25 (2022).
15. Sharp, T.H. et al. Insights into IgM-mediated complement activation based on in situ structures of IgM-C1-C4b. *Proc. Natl. Acad. Sci.* **116**, 11900-11905 (2019).
16. Kumar, N., Arthur, C.P., Ciferri, C. & Matsumoto, M.L. Structure of the human secretory immunoglobulin M core. *Structure* **29**, 564-571.e563 (2021).

17. Chen, Q., Menon, R., Calder, L.J., Tolar, P. & Rosenthal, P.B. Cryomicroscopy reveals the structural basis for a flexible hinge motion in the immunoglobulin M pentamer. *Nature Communications* **13**, 6314 (2022).
18. Hiramoto, E. et al. The IgM pentamer is an asymmetric pentagon with an open groove that binds the AIM protein. *Science Advances* **4**, eaau1199 (2018).
19. Li, Y. et al. Structural insights into immunoglobulin M. *Science* **367**, 1014-1017 (2020).
20. Sarrias, M.R. et al. Biochemical characterization of recombinant and circulating human Spα. *Tissue Antigens* **63**, 335-344 (2004).
21. Gebe, J.A. et al. Molecular cloning, mapping to human chromosome 1 q21-q23, and cell binding characteristics of Spalpha, a new member of the scavenger receptor cysteine-rich (SRCR) family of proteins. *J. Biol. Chem.* **272**, 6151-6158 (1997).
22. Arai, S. et al. Obesity-Associated Autoantibody Production Requires AIM to Retain the Immunoglobulin M Immune Complex on Follicular Dendritic Cells. *Cell Rep.* **3**, 1187-1198 (2013).
23. Gao, X. et al. Assessment of Apoptosis Inhibitor of Macrophage/CD5L as a Biomarker to Predict Mortality in the Critically Ill With Sepsis. *Chest* **156**, 696-705 (2019).
24. Lai, X., Xiang, Y., Zou, L., Li, Y. & Zhang, L. Elevation of serum CD5L concentration is correlated with disease activity in patients with systemic lupus erythematosus. *Int. Immunopharmacol.* **63**, 311-316 (2018).
25. Sanjurjo, L., Aran, G., Roher, N., Valledor, A.F. & Sarrias, M.-R. AIM/CD5L: a key protein in the control of immune homeostasis and inflammatory disease. *Journal of Leukocyte Biology* **98**, 173-184 (2015).
26. Sanchez-Moral, L. et al. Multifaceted Roles of CD5L in Infectious and Sterile Inflammation. *Int. J. Mol. Sci.* **22**, 4076 (2021).
27. Wang, C. et al. CD5L/AIM Regulates Lipid Biosynthesis and Restrains Th17 Cell Pathogenicity. *Cell* **163**, 1413-1427 (2015).
28. Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **17**, 395-398 (2020).
29. Young, G. et al. Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423-427 (2018).
30. Miyazaki, T., Hirokami, Y., Matsuhashi, N., Takatsuka, H. & Naito, M. Increased Susceptibility of Thymocytes to Apoptosis in Mice Lacking AIM, a Novel Murine Macrophage-derived Soluble Factor Belonging to the Scavenger Receptor Cysteine-rich Domain Superfamily. *J. Exp. Med.* **189**, 413-422 (1999).
31. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
32. Fong, R. et al. Structural investigation of human S. aureus-targeting antibodies that bind wall teichoic acid. *MAbs* **10**, 979-991 (2018).
33. Völlmy, F. et al. A serum proteome signature to predict mortality in severe COVID-19 patients. *Life Science Alliance* **4**, e202101099 (2021).
34. Demichev, V. et al. A time-resolved proteomic and prognostic map of COVID-19. *Cell Systems* **12**, 780-794.e787 (2021).
35. Geyer, P.E. et al. High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion. *EMBO Mol. Med.* **13**, e14167 (2021).
36. Sugisawa, R. et al. Independent modes of disease repair by AIM protein distinguished in AIM-felinized mice. *Sci. Rep.* **8**, 13157 (2018).
37. Aran, G. et al. CD5L is upregulated in hepatocellular carcinoma and promotes liver cancer cell proliferation and antiapoptotic responses by binding to HSPA5 (GRP78). *The FASEB Journal* **32**, 3878-3891 (2018).
38. Yamazaki, T. et al. A proteolytic modification of AIM promotes its renal excretion. *Sci. Rep.* **6**, 38762 (2016).
39. Koyama, N. et al. Activation of apoptosis inhibitor of macrophage is a sensitive diagnostic marker for NASH-associated hepatocellular carcinoma. *J. Gastroenterol.* **53**, 770-779 (2018).
40. Li, M. et al. CD5L deficiency attenuate acetaminophen-induced liver damage in mice via regulation of JNK and ERK signaling pathway. *Cell Death Discovery* **7**, 342 (2021).
41. Balakrishnan, L. et al. Differential proteomic analysis of synovial fluid from rheumatoid arthritis and osteoarthritis patients. *Clin. Proteomics* **11**, 1 (2014).
42. Tarazón, E. et al. Plasma CD5L and non-invasive diagnosis of acute heart rejection. *The Journal of Heart and Lung Transplantation* **39**, 257-266 (2020).

43. van der Neut Kolfschoten, M. et al. Anti-Inflammatory Activity of Human IgG4 Antibodies by Dynamic Fab Arm Exchange. *Science* **317**, 1554-1557 (2007).
44. Dingess, K.A. et al. Identification of common and distinct origins of human serum and breastmilk IgA1 by mass spectrometry-based clonal profiling. *Cell. Mol. Immunol.* **20**, 26-37 (2023).

Supporting Information

Supplementary Methods

Serum, saliva, and breastmilk samples

Serum samples were acquired from in-house healthy volunteers who signed written consent or were leftovers from routine diagnostic screening of Tetanus toxoid-boosted donors. Therefore, no informed consent was obtained for the latter samples. Ethical approval was obtained from the Sanquin Ethical Advisory Board. Saliva samples were collected anonymously from healthy volunteers. Human milk was collected by lactating women[1]. Ethical approval was obtained from the Medical Ethics Committee of the Amsterdam UMC, location VUmc and written informed consent was obtained from all participants. All materials were used anonymously without any connection to clinical or person-specific data.

Bottom-up proteomics sample preparation

All samples, including purified human IgM ± CD5L, as well as whole serum (1 µL), saliva (10 µL), or milk (6 µL) were diluted in a buffer 1% (w/v) sodium deoxycholate (SDC), 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 40 mM chloroacetamide (CAA), and 100 mM Tris-HCl at pH 8.0 supplemented with protease inhibitor (cOmplete mini EDTA-free, Roche). Samples were subjected to denaturation and alkylation at 95 °C for 10 min, followed by 4-fold dilution with 50 mM Tris-HCl pH 8.5. Proteases were added at an enzyme-to-protein ratio (w/w) of 1:50 for trypsin (Sigma-Aldrich) and 1:75 for Lys-C (Wako), followed by overnight digestion at 37 °C. Samples were acidified to roughly 1% formic acid (v/v) to precipitate the SDC, followed by centrifugation cycles for 10 min at 20,000 g whereby the supernatant was collected. Samples were stored at -20 °C prior to LC-MS analysis.

Proteome profiling of human biofluids

Samples were loaded onto Evosep® Pure tips, followed by the regular Evosep protocol except for the washing step, which was substituted by two rounds of 50 µL 0.1% formic acid. Chromatographic separation was performed using the 30 SPD method on an Evosep (Evosep) fitted with an EV-1109 analytical column (Dr Maisch C18 AQ, 1.5µm beads, 150 µm ID, 8 cm length, Evosep), using MilliQ water with 0.1% formic acid (v/v) as solvent A and acetonitrile with 0.1% formic acid (v/v) as solvent B. Mass spectrometric analysis was performed on an Orbitrap Exploris 480 mass spectrometer (Thermo Scientific) using a data-independent (DIA) acquisition method. In each cycle, first, an MS1 scan was acquired at a set resolution of 60,000, followed by 40 sequential quadrupole isolation windows of 15 *m/z* covering *m/z* 400-1,000 for HCD MS2 at a set resolution of 15,000. The automatic gain control (AGC) target was set to 300% for MS1 and 1000% for MS2. The maximum injection time was set to 120 ms for MS1 and to "Auto" for MS2. Raw files were searched using DIA-NN (version 1.8) without spectral libraries and with the "Deep learning" option enabled[2]. Trypsin was set as the protease with a maximum of 1 missed cleavage enabled. Cysteine carbamidomethylation was set as a fixed modification, and N-terminal methionine excision was enabled. The precursor and protein false discovery rates were set to 1%. Normalization was performed per sample type (Milk, Serum, Saliva) using the

MaxLFQ algorithm incorporated into DIA-NN. All other settings were set to their default. No imputation was performed.

Proteomics analysis of purified human proteins

Peptide samples of about 100 ng were separated and analyzed by an UltiMate 3000 UHPLC system (ThermoFisher Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific). Peptides were first trapped on a 300 µm x 5 mm trap column packed with C18 PepMap100, 5 µm, 100 Å (Thermo Fisher Scientific, P/N 160454) and then separated on a 75 µm × 500 mm analytical column packed with Poroshell 120 EC-C18, 2.7 µm (ZORBAX Chromatographic Packing, Agilent Technologies). Mobile phase A consisted of 0.1% FA (v/v) in MilliQ water, while mobile phase B consisted of 0.1% FA (v/v) in acetonitrile. Peptides were separated at a 300 nL/min flow rate during a 55 min method as follows: 1 min 9% solvent B; a 1 min ramp from 9% to 13% solvent B; a 35 min separation gradient from 13% to 44% solvent B; a 3 min ramp from 44% to 95% solvent B; 4 min 95% solvent B; a 1 min ramp from 95% to 9% solvent B; 10 min 9% solvent B. Mass spectrometric analysis was performed using a data-dependent acquisition (DDA) method. In each cycle, an MS1 scan was required at a set resolution of 60,000, followed by an HCD MS2 scan at a set resolution of 15,000, both using the "Standard" AGC target and "Auto" maximum injection time settings. Data were searched against the UniProtKB/Swiss-Prot human proteome sequence database with MaxQuant version 1.6.17.0[3] using standard settings. After filtering for proteins with a Q-value ≤ 0.02 and a Score ≥ 50, label-free quantitation (LFQ) values were used to calculate relative abundances.

Complex-centric proteome profiling by SEC-MS

Complex-centric proteome profiling was performed through proteomics analyses on plasma or milk fractions separated by size-exclusion chromatography (SEC). Chromatography was performed using an Agilent 1290 Infinity HPLC system (Agilent Technologies) consisting of a vacuum degasser, refrigerated autosampler with a 100 µL injector loop, binary pump, thermostated two-column compartment, refrigerated fraction collection module, and multi-wavelength detector. Samples were separated using a dual-column setup comprised of a Yarra SEC-4000 and SEC-3000 column (300 x 7.8 mm i.d., 3 µm, 500 Å or 290 Å respectively) (Phenomenex). Separation was performed at 17 °C with a flow rate of 0.5 mL/min using 150 mM aqueous ammonium acetate with 50 mM L-arginine at pH 7.5 as the mobile phase. A volume of 20 µL pooled serum (ca. 1.2 mg of protein) was injected, followed by separation in 60 min, whereby proteins eluted in a 20-42 min window and were collected in 74 fractions. These fractions were subjected to in-solution digestion and proteomics analysis as described above.

Production of recombinant proteins

Recombinant IgM-Fc, IgM anti-biotin, anti-CCP (2D5) and anti-WTA were produced as described previously[4,5]. To produce recombinant human CD5L, FcµR, FcαµR and pIgR, DNA3.1+ expression vectors containing the complete protein sequence (CD5L) or sequence for the extracellular domains (receptors) were designed. All vectors were created with a C-terminal BirA tag (sequence: GLNDIFEAQKIEW[6]) and 10xHis-tag, with the exception of the CD5L vector that was tagged N-terminally instead. The vectors were transfected into HEK 293F cells using PEI-MAX according to the manufacturer's

instructions (Invitrogen). After being cultured for 5 days at 37 °C, 8% CO_2 and shaking at 125 RPM, the supernatant was harvested and filtered through a filter with a pore size of 0.20 µm (Whatman Puradisc 30; Sigma-Aldrich). The recombinant proteins were purified using a HisTrap™ column (Cytiva). The culture supernatant was first dialyzed against phosphate-buffered saline (PBS) and subsequently against a binding buffer (20 mM trisodium phosphate, 0.5 mM sodium chloride, 30 mM imidazole, pH 7.8). The column was washed with water and then binding buffer, after which the supernatant was loaded onto the column. Unbound protein was washed away with binding buffer. The bound proteins were eluted with a gradient of elution buffer (20 mM trisodium phosphate, 0.5 mM sodium chloride, 500 mM imidazole, pH 7.4) and subsequently rebuffered to PBS using a 10 kDa spin column (Amicon Utra-4 Centrifugal Filter Unit; Merck). The concentrations of the purified proteins were determined by measuring the absorbance at 280 nm (NanoDrop One; Thermo Fischer Scientific) after which the samples were stored at -20 °C. The purified recombinant proteins were visualized on SDS page using 4-12% Bis-Tris Protein Gels (NuPAGE™; Invitrogen) according to the manufacturer's protocol, after which they were stained with coomassie (InstantBlue; Expedeon) for at least 30 min or used for western blot. Proteins were site-specifically biotinylated using the Enzymatic Protein Biotinylation Kit (Sigma Aldrich) according to protocol.

Alternatively, IgM-J-CD5L complexes were produced by co-cultures of HEK 293F cells transiently expressing anti-WTA IgM-J or CD5L. Anti-WTA (4497). IgM-J was transfected as described previously[5], whereas separately cells were transfected with pcDNA34 expression vectors encoding CD5L under control of the UPE (Cystatin-S) promoter. For the latter, 0.5 µg total plasmid (50% CD5L plasmid/50% empty vector) was transfected per mL of cell culture. After 5 h of incubation, IgM-transfected and CD5L-transfected cells were combined in a 1 to 1 ratio and 1 mM valproic acid was added to the culture. The produced IgM was purified as described previously[5].

Affinity purification of serum IgM and CD5L

Serum or plasma was diluted 1:1 with PBS supplemented with 500 mM NaCl (mobile phase A), after which samples were loaded on a 10 mL HiTrap column (Cytiva) filled with POROS CaptureSelect IgM affinity matrix (Thermo Fisher) at a flow rate of 1 mL/min using an ÄKTA pure™ 25 (Cytiva). The column was washed with mobile phase A until UV absorption was stable, followed by elution using 100 mM glycine with 500 mM NaCl at pH 3.0 (mobile phase B). Samples were dialyzed overnight against mobile phase A or 20 mM NaAc, 300 mM NaCl (pH 5.5), after which protein concentrations were determined using a NanoDrop (Thermo Fisher Scientific). Samples were stored at -20 °C or -80 °C before further use. Between purifications, the column was cleaned with both 50 mM citric acid pH 2.0 and 6 M guanidine to prevent carryover. Serum CD5L was obtained by reduction of purified serum IgM, with which it co-purifies. Serum IgM was reduced with 1 mM Dithiothreitol (DTT; Calbiogen) for 2 h at 37 °C, followed by size-exclusion chromatographic separation using a Superdex 200 column (10/300 GL; Cytiva). The fractions containing free CD5L were pooled and dialyzed against PBS and stored at -20 °C.

Generation of monoclonal antibodies specific for CD5L

Monoclonal antibodies against CD5L were obtained as described previously[7]. In short, a rabbit was immunized with recombinant CD5L and boosted in 4-week intervals. Whole

blood was obtained 9 days after the third booster, from which PBMCs were isolated and serum was tested for the presence of CD5L antibodies. B cells were isolated from PBMC samples by FACS sorting. Cells were stained with FITC-labeled mouse anti-rabbit IgG (2A9, Abcam) and with CD5L-bt followed by streptavidin-APC, after which the double-positive were sorted and seeded in single wells. These cells were cultured for 9 days, after which the supernatants were tested by ELISA for specific antibodies against free and IgM-bound CD5L.

RNA was isolated of several positive clones, after which cDNA was synthesized, and IGHV and IGLV regions were amplified using PCR[7]. PCR products were sequenced by Sanger sequencing. Antibodies were expressed using synthetic DNA vectors (Invitrogen) coding the variable regions of the antibodies in combination with mouse IgG1 kappa constant domains, essentially as described before[7]. Antibodies were produced by transient transfection using the Freestyle HEK293 system, and antibodies were purified as described previously[4] using a HiTrap protG column (GE Healthcare) and stored in 5 mM NaAc (pH 4.5).

Mass photometry

Mass photometry (MP) experiments were performed on a Refeyn One^{MP} or Samux^{MP} mass photometer (Refeyn). Microscope coverslips (24 mm × 50 mm; Marienfeld) were cleaned by sonication in two sequential iterations between isopropanol and MilliQ water, followed by placement of a CultureWell gasket (Grace Biolabs). Typically, 15 µL of PBS was placed in a well for focusing, after which about 3 µL of diluted sample was introduced and mixed. For IgM, the measurement concentration was typically around 20 nM of monomeric subunits. Measurements were recorded for 120 s using medium field-of-view settings on the One^{MP} or standard settings on the Samux^{MP}. Calibration of the One^{MP} was performed using an in-house mix consisting of IgG4Δhinge-L368A, IgG1-Campath, apoferritin, and GroEL (73, 149, 479, and 800 kDa). Calibration of the Samux^{MP} was performed using thyroglobulin oligomers (670, 1340, and 2010 kDa). Data were processed in DiscoverMP (Refeyn), followed by analysis and plotting using an in-house Python library.

Native Orbitrap-based charge detection mass spectrometry

Orbitrap-based CD-MS was performed on an Orbitrap Q Exactive Plus UHMR instrument (Thermo Fisher Scientific). Firstly, proteins were buffer exchanged to 150 mM aqueous ammonium acetate pH 7.5 through six consecutive dilution and concentration steps at 4 °C using Amicon Ultra centrifugal filters with a 10 kDa molecular weight cutoff (Merck). For experiments studying binding, complexes were assembled by mixing the subcomponents at the desired molar ratios, followed by incubation at room temperature (RT) for at least 30 min. Samples were further diluted using 150 mM aqueous ammonium acetate pH 7.5 to reach the single particle regime, and immediately loaded into gold-coated borosilicate capillaries (prepared in-house) for direct infusion from a static nano-electrospray ionization source. Measurements were performed at low pressure settings using xenon as collision gas at a set resolution of 200,000 at 400 *m/z*, corresponding to a transient time of 1024 ms. The number of microscans was set to 1, noise threshold set to 0, and no transient averaging was employed. Ion injection times and transmission parameters were manually optimized for each measurement to maintain single ion detection. After acquisition, .raw files were processed and filtered as described by Wörner

et al.[8] From the resulting two-dimensional data, ion events corresponding to species of interest were extracted through density-based spatial clustering of applications with noise (DBSCAN), allowing them to be indicated in a separate color for presentation in mass density figures.

B cell cultures

B cell cultures were performed analogously as described in[9]. Buffy coats were obtained from anonymized adult healthy donors upon written informed consent in accordance with the guidelines established by the Sanquin Medical Ethical Committee and in line with the Declaration of Helsinki. Briefly, peripheral blood mononucleated cells were isolated and $CD19^+$ cells were isolated. For T cell dependent (TD) conditions, $CD27^+IgD^-$ (memory) and $CD27^-IgD^+$ (naïve) B cell subsets sorted cells were cultured on a monolayer of irradiated 3T3 fibroblast expressing human CD40L[10] and in B cell medium supplemented with recombinant human IL-21 (50 ng/mL, Preprotech) or IL-21 and IL-4 (50 ng/mL and 25 ng/mL, Preprotech). For T cell independent (TI) culture, cells were instead supplemented with 0.1 μM CpG (Invivogen). After ten (TD) and seven (TI) days of culture, supernatants were analyzed by ELISA (see below).

IgM-CD5L complex ELISA

To measure IgM-CD5L complexes, we set up a sandwich ELISA that uses anti-IgM (MH-15-1; Sanquin) for capture, and anti-CD5L (clone 5B5) for detection. 2 μg/mL a-IgM in PBS was coated on maxisorp plates (Thermo Fisher Scientific) overnight at 4 °C. After coating, plates were washed five times with PBS supplemented with 0.02% Tween-20 (PBST). Samples were diluted in high-performance ELISA buffer (HPE; Sanquin) to the desired concentration and 100 μL of sample was transferred to each well, after which plates were incubated at RT for an hour while shaking at 300 rpm. A pooled reference serum with a known concentration of IgM was used as a calibrator. After sample incubation, the plates were washed and 100 μL of a-CD5L 5B5-bt diluted in HPE (0.5 μg/mL) was transferred to each well. Clone 10D11 or 7E4 were used in a similar setup to check for proper integration of CD5L into recombinant IgM, as these bind to an epitope on CD5L that is shielded when bound to IgM. Plates were then incubated for 1 h at RT while shaking and then washed again. Subsequently, bound a-CD5L was detected with streptavidin-HRP in HPE (1:1000). Alternatively, in order to determine the total amount of IgM in the tested samples, the IgM complexes were instead detected with a-IgM-HRP (0.33 μg/mL in HPE). Both conjugates were incubated at RT for 30 min while shaking. The detection was visualized, and the absorbance was read at 450 nm and 540 nm for background correction. Where appropriate, recombinant CD5L-IgM complexes were used as a calibrator (see below).

Free CD5L ELISA

To measure CD5L not in complex with IgM, we set up a sandwich ELISA using anti-CD5L clones 10D11 and 7E4 that both only bind to free CD5L but recognize different epitopes. 1 μg/mL anti-CD5L 10D11 was coated on maxisorp plates (Thermo Fisher Scientific) overnight at 4 °C. Plates were subsequently washed five times with PBST, after which samples (100 μL) were incubated at RT (1 hr, shaking 300 rpm) after being serially diluted HPE. After sample incubation, the plates were washed and 100 μL of a-CD5L 7E4-bt diluted in HPE (0.5 μg/mL) was transferred to each well. Plates were further developed as

described above for the complex ELISA. Purified recombinant CD5L was used as a calibrator.

Formation and characterization of CD5L-IgM complexes

Recombinant IgM-J or IgM-Fc-J was combined with an excess of recombinant or serum-derived CD5L in a buffer of PBS supplemented with 0.01% Tween-20, 0.1 mM glutathione (GSH; Sigma) and 0.1 mM glutathione disulfide (GSSG; Sigma) and incubated at RT overnight. After incubation, the complexes were further diluted in PBST to the desired concentration. Formed complexes were tested in the IgM-CD5L complex ELISA and a purified serum IgM sample was included as a reference. To further characterize the formed CD5L-IgM complexes, 20 µg of sample in PBS was fractionated using HP-SEC Agilent 1260 Infinity II (Agilent Technologies) with a Superose® 6 Increase 10/300 GL Column (GE Healthcare). Elution peaks were measured at 280 nm absorbance and the size of these peaks was estimated using multi-angle light scattering (MALS) as described previously[4].

Receptor binding to IgM-CD5L complexes

To determine the impact of CD5L integration of binding to IgM receptors, the binding of these complexes was assessed in ELISA. To this end, 100 µL of biotinylated recombinant FcµR or FcαµR at 1 µg/mL in PBST was added to streptavidin-coated plates (Thermo Fischer Scientific) and incubated for 1 h at RT. Plates were washed, after which 100 µL diluted IgM-CD5L complexes diluted in PBS supplemented with 0.2% w/v gelatin (Merck) and 0.1% v/v Tween-20 (PTG) were added to the plate. Plates were incubated for an hour at RT while shaking and then washed. Bound IgM was detected with anti-IgM-HRP (0,33 µg/mL in PTG) for 30 min at RT while shaking. Alternatively, for the reversed pIgR binding ELISA, a-IgM was coated at 2 µg/mL in PBS on maxisorp plates, overnight at 4 °C. Complexes were added in PTG and incubated for an hour while shaking at RT, after which the plates were washed and biotinylated pIgR (1 µg/mL in PTG) was added to the wells. The plates were incubated at RT for an hour while shaking, then washed and the bound pIgR was detected with HRP-labelled streptavidin (1:1000 in PTG) for 20 min at RT while shaking. Detection was visualized and subsequent measurement of absorbance was performed as described above.

Complement assays

Complement activation was assessed in a C3b deposition ELISA and complement-dependent cytotoxicity (CDC) assay as described previously[4]. The C3b deposition ELISA was performed using human serum albumin (HSA; albuman; Sanquin) that was reacted with 30-120 µM biotinylation reagent (EZ-Link™ Sulfo-NHS-LC-Biotin; Thermo Fischer Scientific), whereas for the CDC assay red blood cells were biotinylated at a concentration of 50-500 µM.

For the bacterial killing experiments, *Staphylococcus aureus* Wood 46 was cultured from glycerol stock onto Trypticase Soy Agar II with 5% Sheep Blood plates (254087, BD). A single bacterial colony was picked from a plate and grown overnight in Todd Hewitt broth shaking at 37 °C. Bacteria were subcultured into fresh Todd Hewitt broth and grown until the mid-log phase (OD_{600} = 0.4-0.6). The bacteria were pelleted by centrifugation and

resuspended in RPMI 1640 medium (Thermofisher/Gibco) with 0.05% human serum albumin (RPMI-HSA) to OD_{600} = 1.

To quantify antibody binding, bacteria were diluted to OD_{600} = 0.01 and incubated with a serial dilution of monoclonal anti-WTA (4497) IgM+J ±CD5L antibodies, for 30 min at 4 °C under shaking conditions. Subsequently, the bacteria were washed with RPMI-HSA and resuspended in 3 µg/mL Goat-anti-human-kappa-AF488 (2060-30, Southern Biotech) for 30 min at 4 °C under shaking conditions.

To quantify C3b deposition, bacteria were diluted to OD_{600} = 0.02 and incubated with a serial dilution of monoclonal anti-WTA (4497) IgM+J ± CD5L antibodies for 15 min at 4 °C under shaking conditions. The bacteria were further diluted to OD_{600} = 0.01 and incubated with 1% ΔIgG/IgM-serum as a complement source for 30 min at 37 °C under shaking conditions. ΔIgG/IgM-serum was prepared by depleting human pooled serum from IgG and IgM by affinity chromatography as previously described[11]. After serum incubation, the bacteria were washed with RPMI-HSA, pelleted by centrifugation, and incubated with 3 µg/mL monoclonal mouse anti-C3b (clone bH6), which was randomly labelled with NHS-Alexa Fluor 488 (AF488, Thermo Fisher Scientific) as described previously[12].

After incubation with detection antibodies, the bacteria were washed, pelleted, and resuspended in RPMI-HSA containing 1% PFA as a fixating agent. Surface binding of fluorescent detection antibodies was analyzed by flow cytometry on a MACSQuant VYB (Miltenyi Biotec).

Molecular modeling of IgM-CD5L complex

To model the IgM-CD5L complex, AlphaFold2 multimer[13] was used to predict the structure of CD5L along with the J-chain and two IgM-Fc regions. The CD5L model was then positioned within the previously solved structure of the IgM-Fc pentamer[14] by matching the J-chains from both structures, using the Match Maker tool in UCSF Chimera[15].

Supplementary Figures

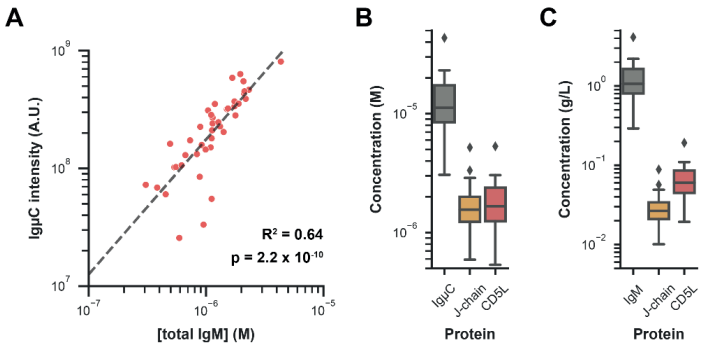


Figure S1 | Absolute quantification of IgμC, J-chain and CD5L in serum. (A) Label-free quantitation intensity values of IgμC correlate well with absolute IgM concentrations determined by ELISA (n = 42). The black line indicates a linear regression model fitted to log-transformed data with R^2 = 0.64 and p = 2.2 x 10^{-10}. To estimate absolute concentrations for all detected proteins by proteomics, for each sample, we first calculated an intensity-to-concentration conversion factor using the IgμC intensity and total IgM concentration. This factor assumed that each IgM molecule holds 10 IgμC chains. Then, the intensity values of all other proteins were converted using this conversion factor. **(B)** Boxplot showing the resulting molar concentrations of IgμC, J-chain, and CD5L in serum, revealing a close to 10:1:1 molecular ratio. **(C)** Boxplot showing the serum concentrations of IgM, J-chain and CD5L in g/L (assuming molecular weights of 950 kDa, 17 kDa and 36 kDa, respectively).

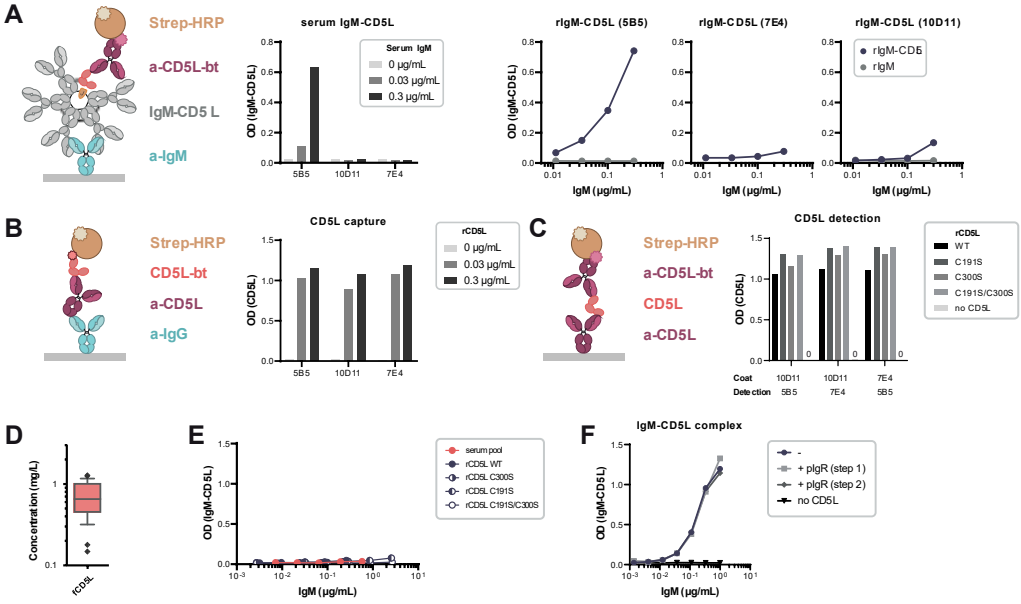


Figure S2 | Characterization of monoclonal anti-CD5L antibodies raised to detect IgM-bound or unbound CD5L, respectively. (A) IgM-bound CD5L was measured in ELISA as depicted, by the capture of IgM by an anti-IgM antibody and detection with one of the anti-CD5L mAbs. Only mAb 5B5 was able to detect serum IgM-bound CD5L (left panel). Similarly, 5B5 detects recombinant IgM-CD5L complexes but does not bind to IgM in the absence of CD5L (right panel). The inability of 10D11 and 7E4 to bind to complexes produced *in vitro* indicates that their respective epitopes of CD5L are shielded in these recombinant complexes as for serum IgM-CD5L. **(B)** CD5L capture by anti-CD5L mAbs was tested in ELISA as depicted. All clones captured CD5L to a similar extent. **(C)** The three anti-CD5L mAbs were tested in combination (*i.e.*, as capture or detection antibody) and bind different epitopes as all combinations gave a signal. Additionally, all clones were able to bind CD5L mutants, showing that detection was not impacted by introduced mutations. **(D)** Free CD5L (fCD5L) was measured in serum (n = 28) using a combination of 7E4 and 10D11 for capture and detection, which both are unable to bind IgM-bound CD5L. Concentrations were quantified with a standard of recombinant CD5L. **(E)** IgM-CD5L complexes formed with WT and recombinant CD5L mutants (Figure 2C) were also assessed in ELISA with mAbs 10D11 and 7E4 as detection. No CD5L was detected, which implies that CD5L C300S is integrated in a manner that shields similar epitopes as WT CD5L. **(F)** Recombinant pIgR was added either during sample incubation (step 1) or detection (step 2) of the IgM-CD5L complex ELISA with clone 5B5 shown in panel A. The presence of pIgR did not affect the binding of 5B5 to the IgM-CD5L complex, which shows that SC does not interfere with IgM-CD5L measurements and interactions with these antibodies.

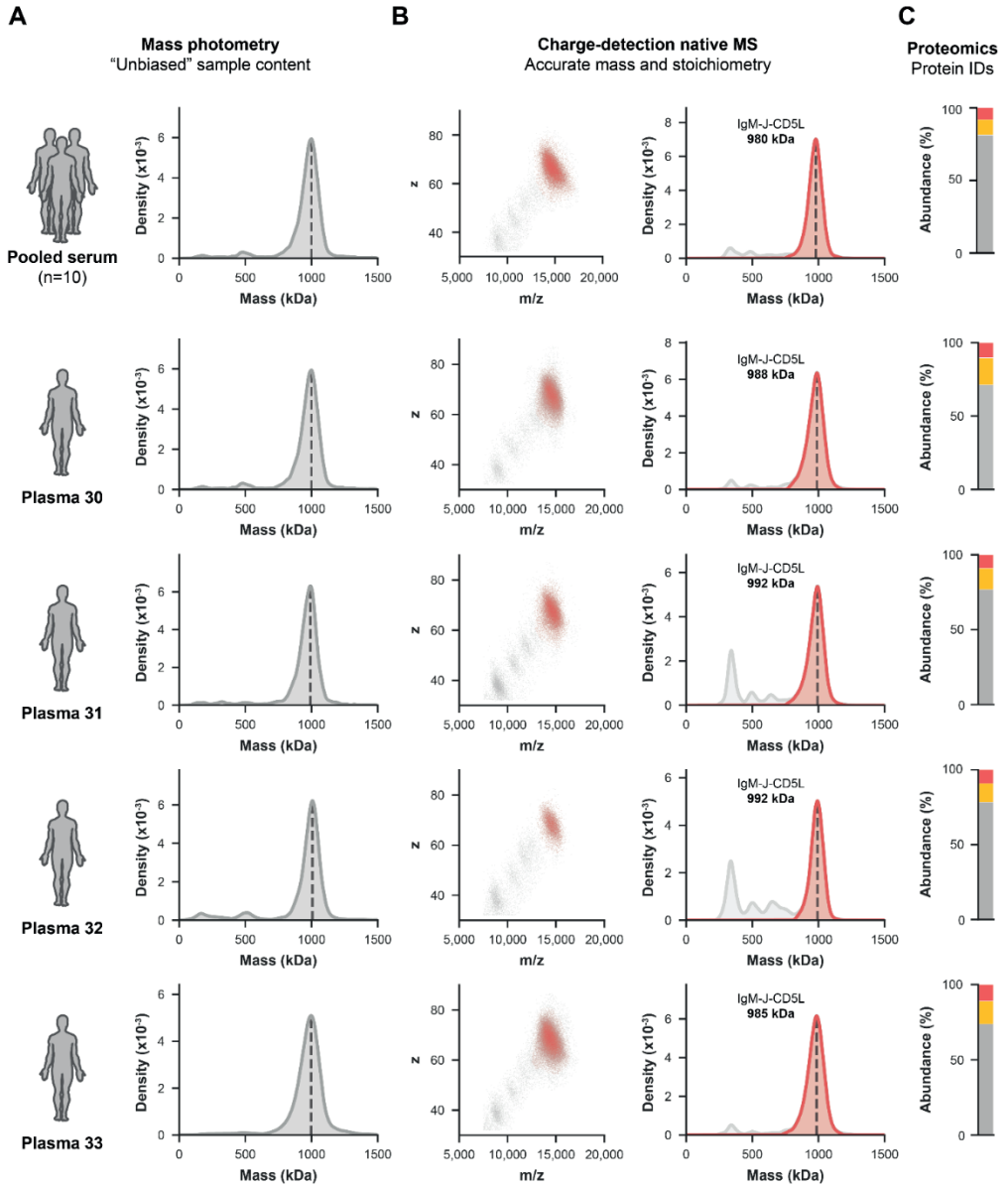


Figure S3 | Circulatory IgM from pooled serum and plasma from individual donors is a pentamer with J-chain and CD5L incorporated. (A) Mass photometry density plots of affinity-purified IgM from pooled serum (n = 10, top row) and plasma from four individual donors (bottom four rows). All samples were highly homogeneous and showed an average IgM mass just below 1 MDa. **(B)** Analysis of the same samples by CD-MS consistently revealed homogenous mass distributions with average masses between 980-992 kDa. This is a mass shift of about + 40 kDa compared to a recombinant IgM-J pentamer ($(IgM)_5$:$(J)_1$, see Figure 1), closely matching the incorporation of one CD5L molecule ($(IgM)_5$:$(J)_1$:$(CD5L)_1$). Shown are the 2D data of *m/z* versus *z* (left), wherein ion events corresponding to IgM were extracted through density-based spatial clustering of applications with noise (DBSCAN), indicated in red. The same measurement is visualized in a 1D mass density plot (right), wherein the extracted IgM ions are overlaid in red against the total measurement in grey. **(C)** The presence of CD5L was confirmed by proteomics, with abundances closely matching a molecular ratio of 1:1:10 of CD5L (red), J-chain (yellow) and IgμC (grey) respectively.

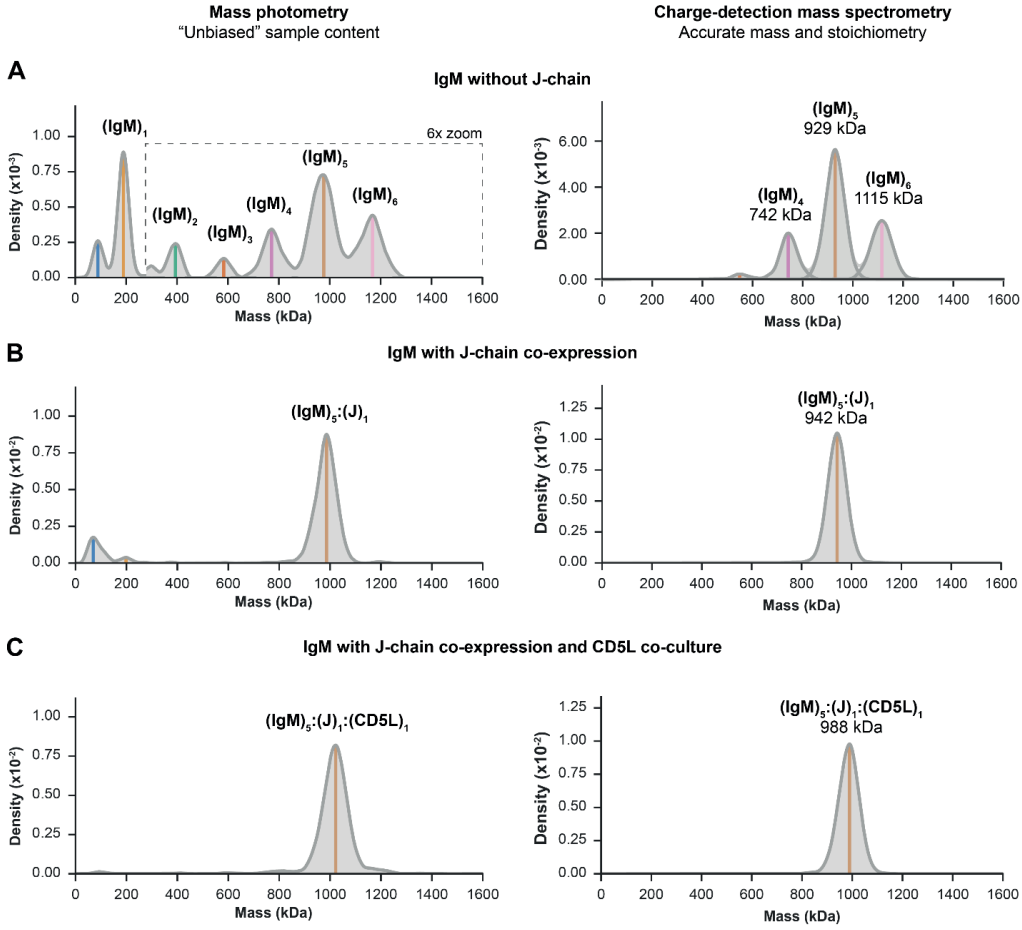


Figure S4 | Characterization of recombinant IgM anti-WTA (4497) without J-chain, with J-chain, and with J-chain and CD5L by MP and CD-MS. (A) IgM without J-chain was produced by co-expression of the HC and LC, resulting in a mixture of oligomers of up to hexamers as measured by MP (left). Accurate mass measurement of larger oligomers by CD-MS confirmed these annotated stoichiometries (right). **(B)** Co-expression of IgM with the J-chain resulted in defined oligomerization into exclusively J-chain-linked pentamers. **(C)** IgM-J-CD5L was produced by coculture of cells co-expressing IgM and J-chain with cells expressing CD5L, which, after careful balancing of the ratio, resulted in near-complete CD5L incorporation and stoichiometrically highly homogeneous samples.

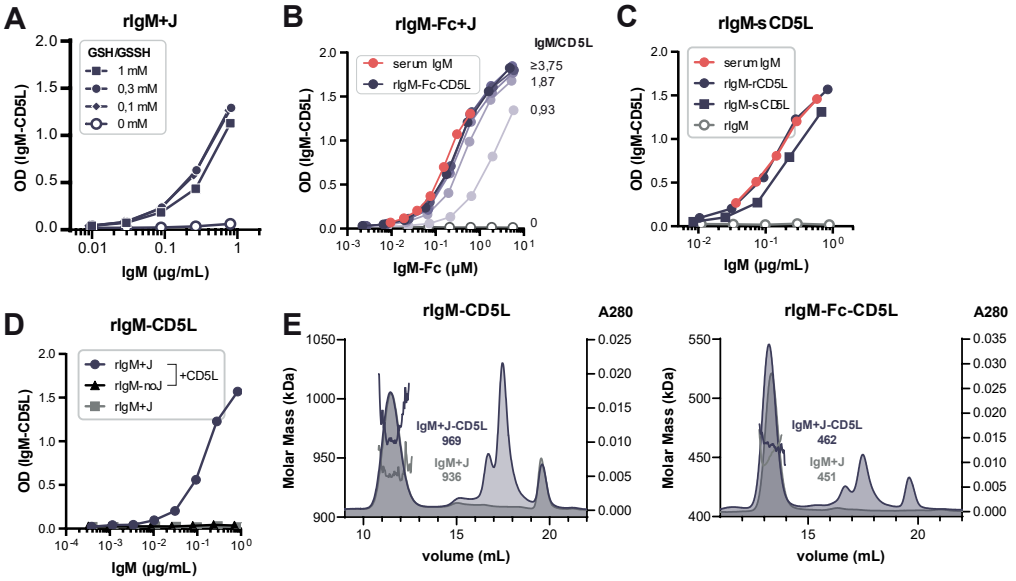


Figure S5 | Assessment of IgM-CD5L complex formation *in vitro*. (A) Recombinant IgM and monomeric free CD5L (molar ratio 1:2) were co-incubated in the presence and absence of a glutathione redox buffer. Only in the latter condition, efficient integration of CD5L into IgM is observed, which implies disulfide bond formation/shuffling is important for complex formation. **(B)** Complex formation can also be achieved with recombinant IgM-Fc (CH2-CH4) and CD5L. **(C)** Serum CD5L (sCD5L) was purified from serum IgM by reduction with 1 mM DTT for two hours and subsequent HP-SEC, after which the free sCD5L-containing fractions were pooled. sCD5L can, similarly to rCD5L, be (re-)integrated into rIgM. **(D)** rCD5L does not bind to recombinant IgM that lacks J-chain and only forms complexes with IgM-J. **(E)** Recombinant IgM-CD5L or IgM-Fc-CD5L complexes were assessed by HP-SEC followed by multi-angle light scattering analysis. Elution profiles (right y-axis) and estimated molecular mass (left y-axis) are shown. The addition of CD5L to IgM results in a mass shift corresponding to the integration of one CD5L molecule per IgM.

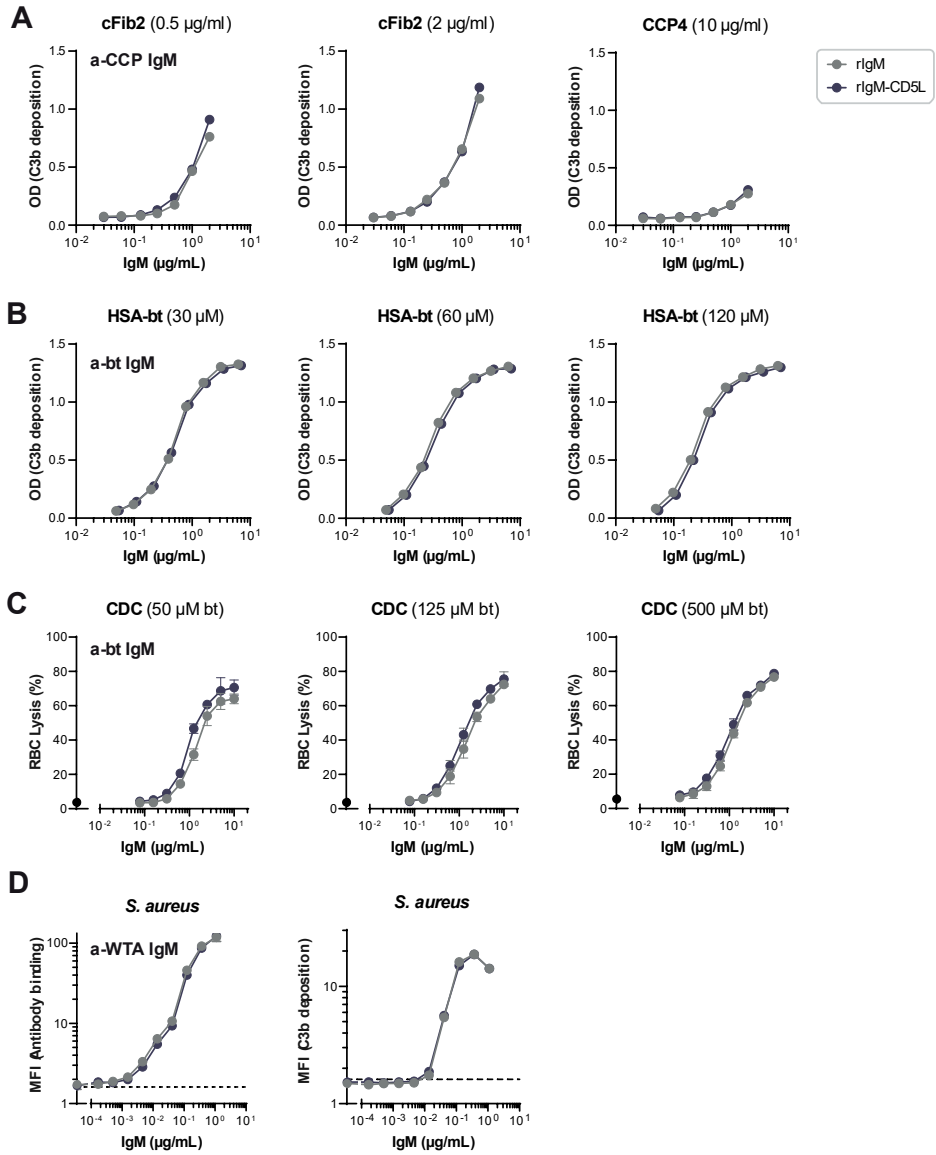


Figure S6 | Complement activation by IgM is not affected by CD5L integration. (A) C3b deposition by an anti-CCP (2D5) or **(B)** anti-biotin antibody upon binding to synthetic citrullinated peptides (cFib2 and CCP4) or biotinylated human serum albumin (HSA-bt), respectively, in the presence of 2.5% human serum. Several antigen densities were tested, ranging from relatively low to fully saturated. Representative plots of n = 3 experiments. **(C)** Complement-mediated lysis of biotinylated red blood cells (RBCs) by an anti-biotin antibody in the presence of 10% human serum. RBCs were biotinylated at different concentrations and subsequently incubated with antibody dilutions and serum for 90 min. Released hemoglobin was determined as a measure of cell lysis. Data are plotted as the mean ± SE of n = 3. **(D)** Antibody binding (left panel) and C3b deposition (right panel) by anti-WTA (4497) antibody on *S. aureus* Wood 46. To study C3b deposition, bacteria were opsonized with IgM for 15 min at 4 °C and then incubated with 1% human serum depleted of IgG and IgM. Binding of antibodies and C3b deposition were determined by flow cytometry. Dotted line represents the detection of the isotype control. Data are plotted as the mean ± SE of n = 3.

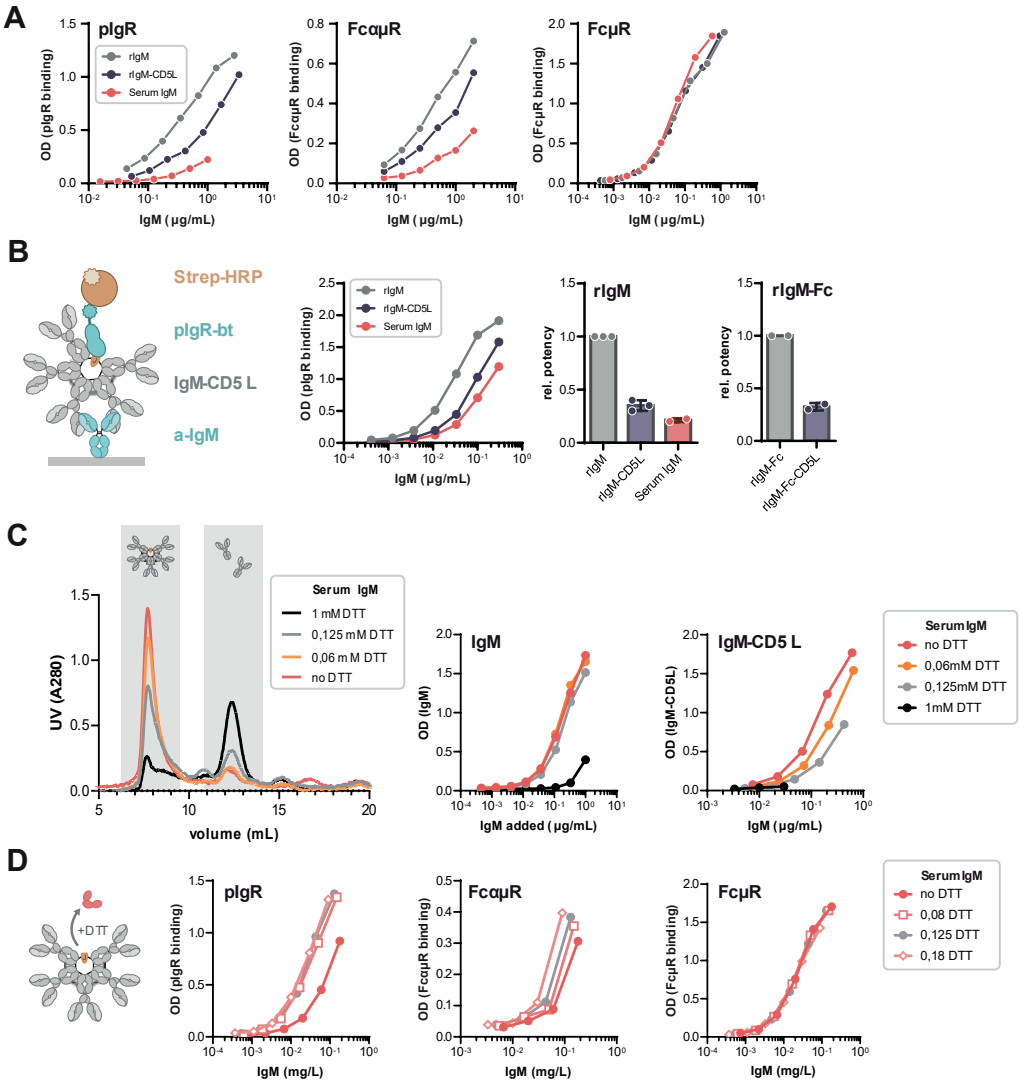


Figure S7 | Binding of IgM to pIgR and FcαμR, but not FcμR, is decreased upon CD5L integration. (A) Representative titration curves for recombinant IgM ± CD5L and serum IgM-CD5L to pIgR, FcαμR and FcμR, of which combined data is shown in Figure 3B. **(B)** Binding of IgM and IgM-Fc (CH2-CH4) ± CD5L to pIgR was also assessed in a reversed setup, where IgM was first captured with an anti-IgM antibody and then binding of biotinylated pIgR was detected with streptavidin-HRP. In both approaches, the integration of CD5L reduced binding to the receptor. Data from n = 2 or 3 experiments. Bars and error bars represent mean and SE. **(C)** Purified serum IgM was reduced using varying concentrations of DTT ranging from 0.06 mM to 1 mM. Reduced IgM was then assessed using HP-SEC and IgM(-CD5L) ELISA. Whereas at high concentrations of DTT IgM falls apart into fragments the size of IgG-like IgM "monomers", at lower concentrations it remains mostly the size consistent with a pentamer. Nevertheless, at these limiting reducing conditions, CD5L is partially released, as assessed by IgM-CD5L complex ELISA (right panel). **(D)** Partial release of CD5L improves binding to pIgR and FcαμR, but not FcμR. Shown is a representative experiment of n = 3.

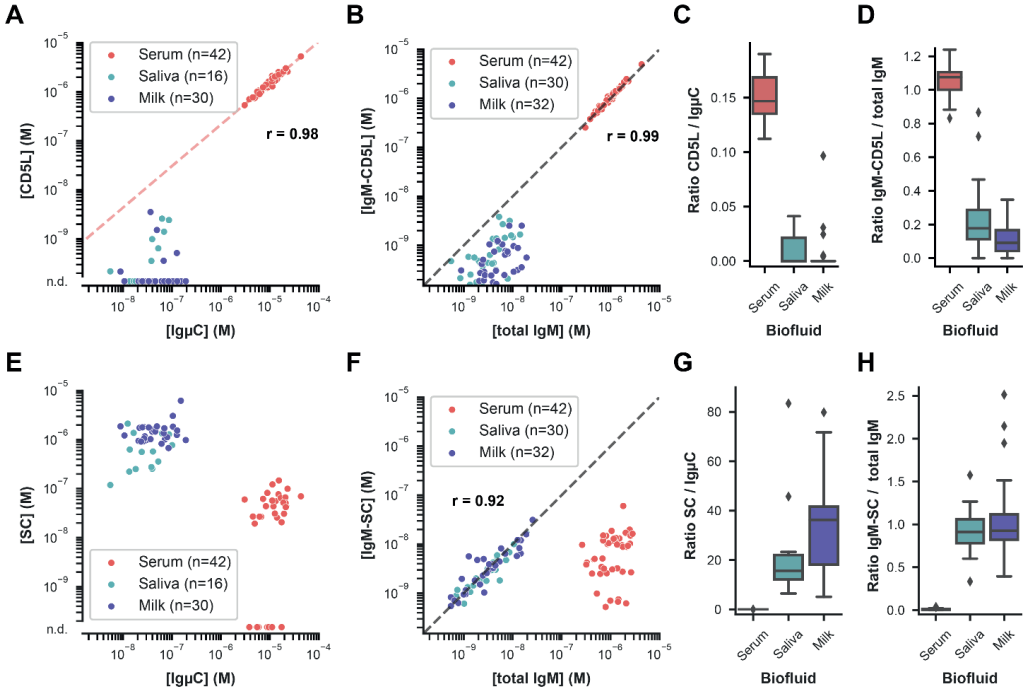


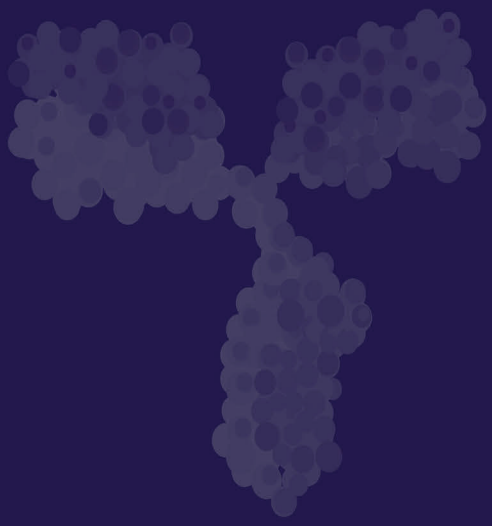
Figure S8 | Secretory IgM is largely devoid of CD5L and contains instead the SC of pIgR. (A) Total levels of CD5L and IgM heavy chain (IgμC) show a remarkably strong correlation in serum (r = 0.98) but not in saliva and milk, as determined by label-free proteomics. CD5L levels in saliva and milk were very low, frequently even below the detection limit. These values are indicated at the bottom as not determined (n.d.). The red line indicates a linear regression model fitted to logarithmically scaled serum data. **(B)** Levels of IgM-CD5L complexes and total IgM similarly show a high correlation in serum (r = 0.99) but not in saliva and milk, as determined by ELISA (see Figure 1). The grey line indicates a 1:1 molecular ratio. **(C)** The molecular ratio between CD5L and IgμC was approximately 1:7 in serum, closely matching the incorporation of one CD5L molecule per IgM pentamer. In saliva and milk, this ratio is much lower. **(D)** Similarly, while the ratio between IgM-bound CD5L and total IgM was approximately 1:1 in serum, IgM from saliva and milk was nearly devoid of CD5L. **(E)** Total levels of SC, the extracellular part of pIgR, were very high in saliva and milk, frequently over an order of magnitude higher than IgμC levels. In serum, SC levels were much lower, sometimes even below the detection limit. These values are indicated as not determined (n.d.). **(F)** Consequently, levels of IgM-SC complexes and total IgM ELISA similarly show a good correlation in saliva and milk (r = 0.92) but not in serum. The grey line indicates a 1:1 molecular ratio. **(G)** The molecular ratio between SC and IgμC was in the order of 10:1 to 40:1 in saliva and milk, highlighting a large excess. In contrast, the ratio in serum was only 0.002:1. **(H)** Whereas the ratio between IgM-bound SC and total IgM was approximately 1:1 in saliva in milk, serum IgM was principally devoid of SC.

Supplementary References

1. Michaud, E., Mastrandrea, C., Rochereau, N. & Paul, S. Human Secretory IgM: An Elusive Player in Mucosal Immunity. *Trends Immunol.* **41**, 141-156 (2020).
2. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41-44 (2020).
3. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).
4. Oskam, N. et al. At Critically Low Antigen Densities, IgM Hexamers Outcompete Both IgM Pentamers and IgG1 for Human Complement Deposition and Complement-Dependent Cytotoxicity. *J. Immunol.* **209**, 16-25 (2022).
5. Aguinagalde, L. et al. Promoting Fc-Fc interactions between anti-capsular antibodies provides strong complement-dependent immune protection against Streptococcus pneumoniae. *bioRxiv*, 2022.2001. 2021.477211 (2022).
6. Fairhead, M. & Howarth, M. in Site-Specific Protein Labeling: Methods and Protocols. (eds. A. Gautier & M.J. Hinner) 171-184 (Springer New York, New York, NY; 2015).
7. Großerichter-Wagener, C. et al. Biased anti-idiotype response in rabbits leads to high-affinity monoclonal antibodies to biologics. *MAbs* **12**, 1814661 (2020).
8. Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **17**, 395-398 (2020).
9. Koers, J. et al. CD45RB Glycosylation and Ig Isotype Define Maturation of Functionally Distinct B Cell Subsets in Human Peripheral Blood. *Front. Immunol.* **13** (2022).
10. Unger, P.-P.A. et al. Minimalistic In Vitro Culture to Drive Human Naive B Cell Differentiation into Antibody-Secreting Cells. *Cells* **10**, 1183 (2021).
11. Zwarthoff, S.A., Magnoni, S., Aerts, P.C., van Kessel, K.P.M. & Rooijakkers, S.H.M. in The Complement System: Innovative Diagnostic and Research Protocols. (ed. L.T. Roumenina) 21-32 (Springer US, New York, NY; 2021).
12. Heesterbeek, D.A. et al. Bacterial killing by complement requires membrane attack complex formation via surface-bound C5 convertases. *The EMBO Journal* **38**, e99852 (2019).
13. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
14. Chen, Q., Menon, R., Calder, L.J., Tolar, P. & Rosenthal, P.B. Cryomicroscopy reveals the structural basis for a flexible hinge motion in the immunoglobulin M pentamer. *Nature Communications* **13**, 6314 (2022).
15. Pettersen, E.F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).

PART 2

TOWARDS HUMAN ANTIBODY DISCOVERY BY MASS SPECTROMETRY



CHAPTER 6

INTRODUCTION TO ANTIBODY DISCOVERY AND THE EMERGING ROLE OF MASS SPECTROMETRY-BASED DE NOVO SEQUENCING THEREIN

In **Part 2** of my thesis, I describe the work that I performed to further develop mass spectrometry-based methodologies for *de novo* sequencing of endogenous human antibodies. **Chapter 6** starts with a brief introduction to the processes by which antibodies are generated and how this knowledge can be used to discover antibodies with therapeutic potential. The next section focuses on the emerging role of mass spectrometry in this process, enabling protein-level analysis of antibody repertoires.

6.1 Discovering Antibodies with Therapeutic Potential from Natural Repertoires

The first therapeutic use of antibodies started almost a century ago in the 1930s when endogenous antibodies were extracted from the blood of healthy donors. In intravenous immunoglobulin therapy (IVIg), such bulk mixtures of antibodies are sometimes still used to treat a variety of conditions, primarily those involving dysregulated antibody production[1]. Modern-day use of antibodies is, however, much more targeted and defined[2]. It typically involves high-quality monoclonal antibodies (mAbs) that bind strongly and selectively to a disease-related antigen, which are manufactured by large-scale recombinant expression. Accordingly, to obtain an antibody of interest, the relevant DNA sequence needs to be uncovered so that it can be used to produce the recombinant version in a (mammalian) cell line. Recombinant production also has the advantage that the antibody can be engineered to enhance certain characteristics, for example by introducing mutations that enhance serum half-life or FcR binding, or by making an originally non-human antibody sequence more human-like[3,4]. Furthermore, additional features can be added that are not found in nature, such as fusion proteins that enhance immune activation or transport to a tissue of interest[5,6]. Unleashing the full power of antibodies for therapeutic use thus critically depends on methods for discovering high-quality sequences. Therefore, it is key to first understand how antibodies are generated by the human body.

Antibodies Are Generated by B Cells

All antibodies originate from a diverse repertoire of B cell receptor (BCR) molecules expressed on the surface of B lymphocytes – or simply B cells. In essence, BCRs are membrane-anchored antibodies that contain an additional disulfide-linked signal transduction moiety, acting as a sensor for antigen binding[7,8]. Naïve B cells are screened for antigen binding by their unique BCRs, which are then optimized and translated into secreted antibodies[9,10]. In this two-step process, the immune system first depends on its ability to generate vastly diverse naïve BCR repertoires with weak but promiscuous binding properties. Secondly, stringent selection and optimization mechanisms are needed to extract relevant BCRs for maturation into more effective antibodies exhibiting high affinity and selectivity.

The incredible diversity of the human naïve BCR repertoire originates from how these proteins are processed at the genomic level[11]. Both HCs and LCs are encoded in four separate genes, namely the variable (V), diversity (D), joining (J), and constant (C) genes, with the LC lacking the D-gene. These genes further exist as multiple alleles, which are assembled in the process of V(D)J recombination[12]. With about 10,000 HC and 350 LC recombinations, the total number of BCR sequences already exceeds 3.5 million at the germline level[13]. Adding to this diversity, the error-prone nature of this process leads to mutations in the variable regions as well as insertions and deletions at the junctions[14]. Especially the CDR3 is heavily affected by this, as it is located at the junction between the V-gene and the D- and/or J-genes. This results in a staggering number of potential naïve

BCR sequences. This vast potential repertoire enables the body to generate antibodies against a vast diversity of antigens.

When naïve B cells in secondary lymphoid organs are presented with an antigen, those that interact are primed and activated to generate the antibody response[9,15]. This directly initiates a first line of defense, which includes the differentiation into both memory B cells to store the naïve sequence as well as short-lived plasma cells that initially produce oligomeric IgM[16]. Simultaneously, a subset of B cells travels to the follicles of the lymph nodes to form germinal centers (GCs), in which they rapidly optimize their BCR in the iterative process of affinity maturation[15,17]. This encompasses the induction of hypervariability in the CDRs by somatic hypermutation in the GC dark zone, combined with positive and negative selection mechanisms mediated by T cells in the GC light zone. Later in the response, class switching and re-entry into the GC dark zone can occur to produce further matured sequences and alternative isoforms (*e.g.*, yielding IgA or IgG)[18]. The highly matured GC B cells typically differentiate into long-lived antibody-secreting plasma cells, whereas those of lower maturity differentiate into memory B cells[19-23]. This discrepancy is thought to enhance the flexibility of the memory B cell pool, enabling quick reactivation and adaptation of the response in a subsequent encounter.

Because memory B cells require antigen presentation for future reactivation, they often localize to strategic sites, such as the spleen and the lymph nodes[24-26]. Smaller portions also enter circulation or reside in mucosal and intestinal tissues[27-29]. The most mature memory B cells and long-lived plasma cells can be found in the bone marrow, however, where they can persist for decades to provide long-term immunity[30-33].

Large Antibody Repertoires Originate from Massive BCR Repertoires

Although the total repertoire of BCR sequences may be enormous, the number of unique antibodies active in circulation is orders of magnitude lower. It has been postulated that the theoretical sequence space for unique human BCRs is beyond 10^{15}, a mind-bogglingly large number[17]. With an estimated total of 1-2 x 10^{11} B cells in the human body, full-scale analysis of BCR repertoires has therefore long seemed prohibitively difficult[34]. Nonetheless, recent advances in high-throughput single-cell genetic sequencing have led to breakthroughs in BCR repertoire analysis. Estimates now range from 10^6 to 10^9 unique BCRs in individuals, though it should be noted that these numbers include both naïve and mature sequences[35-37]. Other studies indicate that perhaps only thousands of unique activated B cells are circulating in the peripheral blood at a time, far fewer than previously thought[37]. Furthermore, theoretical calculations have shown that the total antibody content of blood can only accommodate in the order of 10^4 unique antibodies[38]. This is because binding kinetics require a minimum concentration for antigen binding to occur effectively. The complexity of the serum antibody repertoire may therefore be much lower than that of the initial huge repertoire of BCRs.

Genetic Material is the Standard Source of Antibody Sequence Information

Despite the enormity of the B cell repertoire, the availability of cloning and DNA/RNA sequencing techniques has made their genetic material the typical starting point for antibody discovery[39]. Starting in the 1970s, the first mAbs were obtained by screening B cells from immunized animals. Memory B and plasma cells were typically harvested from the spleen, followed by fusion with a myeloma cell line to create stable hybridomas that express the antibody. After sorting and screening for target binding, the relevant genes could then be cloned into other constructs or be sequenced. This approach is still quite common and is now also applied to mice with humanized immune systems to generate human antibodies that are better tolerated in the clinic[40]. Another established method for antibody discovery utilizes phage display to screen large VH and VL libraries obtained from the mRNA of B cells, often as single-chain variable fragments (scFv). The main advantage of this approach is that libraries of over 10^{10} fragments can be made, even using circulating memory B cells from humans. This negates the need for immunization, as (weak) binders to any target can often be found, and it enables the generation of fully human antibodies. Drawbacks include that original chain pairing information is often lost, and that artificial affinity maturation is required to obtain sequences of sufficient quality.

In recent years, antibody discovery efforts have particularly been boosted by the advent of single-cell sorting and next-generation DNA/RNA sequencing (NGS) techniques[41-45]. Such approaches screen large populations of single B cells for functional properties such as target binding. Mostly memory B cells are used in this case, as their surface BCR expression allows for pre-enrichment. Nonetheless, also plasma cells can be used indirectly through single-cell sorting, expansion, and subsequent screening of secreted antibodies in the culture wells of surviving cells[46,47]. After obtaining B cells of interest, mRNA encoding for the VH and VL regions is extracted and converted into cDNA for sequencing. As processing occurs on the level of single cells, a major advantage is that chain pairing information is retained. Furthermore, with high-throughput sequencing efforts, these approaches now comfortably process many thousands of BCRs per donor, often generating many viable sequence hits. When applied to circulating B cells, single-cell BCR sequencing is also particularly useful for the discovery of human antibodies in donors who have developed immunity to a disease. This has made it possible to discover and elucidate the sequences of neutralizing antibodies from patients recovered from Ebolavirus[48] or SARS-CoV-2[49] infections for example.

6.2 The Emerging Role of Mass Spectrometry in Protein-Level Analysis of Antibody Repertoires and Antibody Discovery

While indirect genomics-based approaches have proven very powerful, the ultimate source of antibody information is of course the protein itself (**Figure 1**). A major constraint of relying on genetic material is that it requires access to the relevant cells. Unfortunately, this is not always straightforward. Most antibodies are produced by long-lived plasma cells that reside in the bone marrow, which are evidently challenging and very invasive to retrieve. When the goal is to analyze human antibody responses, one can

therefore typically only access circulating B cells obtained from blood donations. These cells only make up a small proportion of the total B cell pool, and they are diluted by many immature sequences. Furthermore, not every BCR sequence corresponds to an actually produced antibody, and even those that do, may not need to be expressed at the current moment. The most accessible and relevant source of antibody sequence information is therefore the circulating antibody protein itself[38]. Protein sequencing guarantees researchers that they end up with the fully matured final sequence of the circulating antibody, one that the body deemed worthy of large-scale production. Furthermore, protein-level analysis can uniquely provide quantitative information, which could serve as an additional indicator of biological relevance.

In recent years, several mass spectrometry (MS)-based approaches have been developed for monitoring antibody repertoires and identifying clones of interest. Some of the first of such approaches combined bottom-up MS of HC CDR3 peptides with genomic databases resulting from single-cell NGS of peripheral B cells from the same donors[50-54]. These studies showed how antibody responses are generated in response to infection or vaccination, and they strikingly demonstrated that the peripheral B cell repertoire indeed poorly reflects the serum antibody pool. Alternatively, in LC-MS mass profiling, reduced antibody chains or proteolytically formed Fabs are subjected to chromatographic separation and subsequent intact mass analysis[55-57]. This fully protein-based approach identifies unique clones by their mass and retention time, which can be used to track their abundance across samples. By spiking in a known quantity of mAbs, even absolute concentrations can be calculated at the level of the individual clone. Such analyses have made it possible to investigate the complexity and dynamics of individual antibody repertoires, for example in response to infection or vaccination[56-59]. In contrast to the previously expected diversity of millions of unique clones, these approaches revealed that IgG1 and IgA1 repertoires are dominated by only hundreds of detectible clones that make up 50-90% of the total isotype concentration in blood. Furthermore, clonal concentrations were found to be highly variable, with some diseased donors producing single clones up to the g/L range. Similar experiments can also be performed in combination with antigen-specific enrichment, revealing the polyclonal response against a target of choice, for instance, the SARS-CoV-2 spike protein and variants thereof[60,61]. Intact mass profiling thus presents a practical approach for identifying samples and clones of interest for sequence discovery.

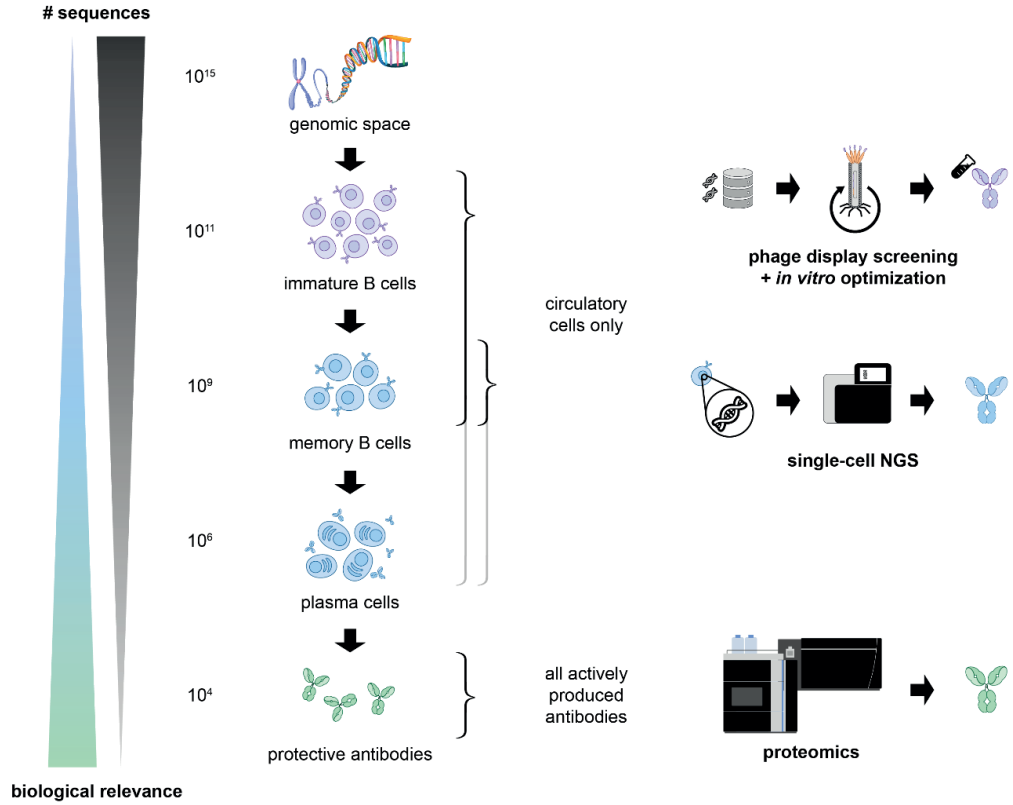


Figure 1 | Techniques for discovering human antibodies and analyzing repertoires. Though very powerful, established methods for obtaining antibody sequences from genetic material of B cells have several drawbacks due to their indirect nature. Firstly, they are limited to circulating B cells, whereas most antibodies are produced by long-lived plasma cells in the bone marrow. When used as a basis for phage display libraries (top), resulting sequences often require artificial *in vitro* affinity maturation. Circulating memory B cells and sometimes plasma cells can also be harvested for single-cell NGS (middle), which has potential for discovering fully human sequences in donors who developed immunity to a disease. However, these sequences do not necessarily correspond to actively produced antibodies, and they are unlikely to be the most matured. Protein-level analysis by MS (bottom), however, enables direct characterization of the antibodies themselves. This guarantees researchers that they end up with the fully matured final product, fully optimized and selected *in vivo*.

Towards MS-Based *De Novo* Sequencing of Endogenous Antibodies

Protein-level sequencing of mAbs is nowadays achievable by bottom-up LC-MS workflows, which analyze peptides resulting from a proteolytic digest (see **Chapter 1**)[62]. In contrast to conventional proteomics experiments where mass spectra are searched against a database, protein sequencing of antibodies requires *de novo* annotation to generate sequence reads (**Figure 2**). Such experiments therefore often use advanced fragmentation schemes such as stepped HCD or EThcD to maximize fragment ion coverage[63,64]. Additionally, experiments using proteases of different specificities are combined to generate overlap between the reads, enabling assembly into full-length protein sequences[65,66]. For highly pure antibody samples, this somewhat laborious approach now

may generate correct annotations of the full polypeptide backbone, also when no genomic data are available[67-70].

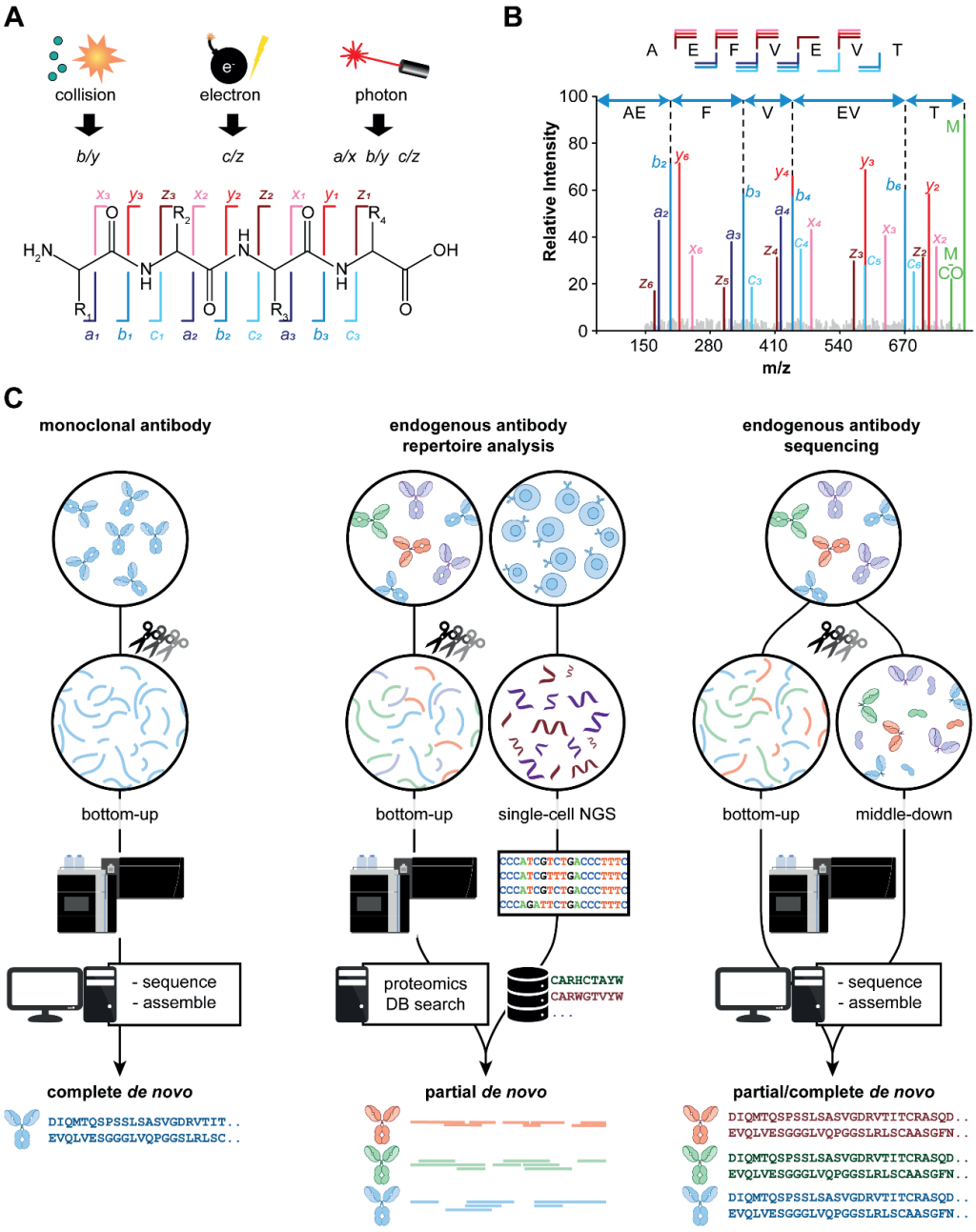


Figure 2 | Overview of MS-based *de novo* sequencing approaches. (A) Different ion activation techniques produce different fragment ion types that can be used to infer amino acid sequences of polypeptide chains. **(B)** By matching mass differences in fragment ion spectra to amino acid residue masses, one can obtain sequence reads. **(C)** For the MS-based sequencing highly purified mAbs, bottom-up approaches using multiple proteases generate sufficient overlapping reads for complete *de novo* sequence assembly. This task

becomes more complex for endogenous antibodies in mixtures. One solution is to match bottom-up data with a database obtained from single-cell NGS of B cells, providing templates for partial *de novo* assembly. Alternatively, peptide-level information can be combined with top- or middle-down MS to provide intact chain masses and sequence tags. Using this information as constraints in sequence assembly enables the complete *de novo* sequencing of endogenous antibodies. Adapted from de Graaf *et al.* (2022)[62].

Although MS-based *de novo* sequencing of mAbs has thus become feasible, the analysis of endogenous antibodies from biofluids is a much more complex challenge. Even when employing antigen-specific capturing, sample complexity rapidly becomes prohibitive for high-confidence assignment of peptide-level data[38]. This is because, owing to the short nature of peptide reads, it quickly becomes elusive which should be assembled into the same polypeptide chain. Additionally, chain pairing information is lost when multiple antibodies are processed together for proteolytic digestion. Proteomics analyses of endogenous antibodies are therefore frequently combined with single-cell NGS experiments to build databases to provide templates for MS-based sequencing[71,72]. However, for purely MS-based analysis of endogenous antibodies, alternative methods and additional layers of information are needed to complement peptide-centric approaches.

Full *de novo* sequencing of endogenous antibodies would benefit greatly from analysis at the intact level, preferably in combination with fragmentation by top- or middle-down MS approaches. The main advantage of these approaches is that complete variable regions and chain pairing can be retained within the same precursor ion, which can be isolated with very high precision. Although the efficient fragmentation and analysis of larger ions is challenging, several advanced ion activation techniques have been pioneered in combination with high-resolution Fourier transform mass spectrometry. Most notable are UVPD[73], ETD[74,75] and ECD[76,77], which are sometimes further enhanced through supplemental activation by CID. Such approaches can yield extensive fragmentation of antibodies, often with multiple co-occurring fragment ion types[78-84]. When the sequence is known *a priori*, this plethora of fragments is often shown to provide high coverage, enabling the identification of mutations and PTMs. The current state-of-the-art, however, does not yet yield complete and unambiguous sequence information for *de novo* interpretation, owing mostly to spectral congestion and signal dilution by multiple fragment ion types. Further development of top- and middle-down MS workflows focusing on reducing spectral complexity could therefore greatly benefit *de novo* sequencing efforts for antibodies. Nonetheless, intact (chain) masses and fragment ion ladders obtained from current methods can already provide additional constraints to facilitate peptide-based sequence assembly. This allows researchers to drastically reduce the number of potential solutions to the assembly problem, which is imperative for the analysis of more complex samples. Very recently, this combination of peptide-centric and protein-centric MS approaches has made it possible to derive full sequences of highly enriched endogenous antibodies, demonstrating proof-of-concept for MS-based antibody discovery[56,69].

6.3 Outline of Part 2

In the following chapters of this thesis, I focus on the further development of new methodologies for MS-based *de novo* sequencing of endogenous antibodies.

In **Chapter 7**, I explore the use of native ECD MS on various IgG subclasses. This enabled me and my coworkers to generate straightforward-to-read sequence ladders of only a single fragment ion type to cover the important CDR3 of both LC and HC, independent of the IgG subclass that was used as a precursor. However, we did observe that the distinct disulfide patterns of the different IgGs influenced the observed fragmentation patterns.

In **Chapters 8 and 9**, we expand this method to heavily glycosylated IgA antibodies and large oligomeric IgM and IgG assemblies, demonstrating similar applicability of the approach, even though these latter molecules are 5-6 times larger in size and more heterogeneous in structure due to the number of occupied glycosylation sites they harbor.

Finally, in **Chapter 10**, we bring antibody mass profiling, top-down, middle-down, and bottom-up MS approaches together to identify and sequence a highly abundant endogenous IgG1 occurring in a patient with a plasma cell disorder.

References

1. Negi, V.-S. et al. Intravenous Immunoglobulin: An Update on the Clinical Use and Mechanisms of Action. *J. Clin. Immunol.* **27**, 233-245 (2007).
2. Lu, R.-M. et al. Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* **27**, 1 (2020).
3. Wang, X., Mathieu, M. & Brezski, R.J. IgG Fc engineering to modulate antibody effector functions. *Protein & Cell* **9**, 63-73 (2018).
4. Chiu, M.L. & Gilliland, G.L. Engineering antibody therapeutics. *Curr. Opin. Struct. Biol.* **38**, 163-173 (2016).
5. Pardridge, W.M. Blood–brain barrier drug delivery of IgG fusion proteins with a transferrin receptor monoclonal antibody. *Expert Opinion on Drug Delivery* **12**, 207-222 (2015).
6. Hutmacher, C. & Neri, D. Antibody-cytokine fusion proteins: Biopharmaceuticals with immunomodulatory properties for cancer therapy. *Adv. Drug Del. Rev.* **141**, 67-91 (2019).
7. Su, Q. et al. Cryo-EM structure of the human IgM B cell receptor. *Science* **377**, 875-880 (2022).
8. Ma, X. et al. Cryo-EM structures of two human B cell receptor isotypes. *Science* **377**, 880-885 (2022).
9. Weaver, C. & Murphy, K. in Janeway's Immunobiology, Edn. 9th 399-444 (Garland Science, 2016).
10. Oliver, B. Generation of Antibody Diversity. (2018).
11. Weaver, C. & Murphy, K. in Janeway's Immunobiology, Edn. 9th 173-212 (Garland Science, 2016).
12. Fanning, L.J., Connor, A.M. & Wu, G.E. Development of the Immunoglobulin Repertoire. *Clinical Immunology and Immunopathology* **79**, 1-14 (1996).
13. Ginestoux, C., Vol. 2023 (1997).
14. Jeske, D.J., Jarvis, J., Milstein, C. & Capra, J.D. Junctional diversity is essential to antibody activity. *J. Immunol.* **133**, 1090-1092 (1984).
15. Akkaya, M., Kwak, K. & Pierce, S.K. B cell memory: building two walls of protection against pathogens. *Nature Reviews Immunology* **20**, 229-238 (2020).
16. Palm, A.-K.E. & Henry, C. Remembrance of Things Past: Long-Term B Cell Memory After Infection and Vaccination. *Front. Immunol.* **10** (2019).
17. Schroeder, H.W. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev. Comp. Immunol.* **30**, 119-135 (2006).
18. McHeyzer-Williams, L.J., Milpied, P.J., Okitsu, S.L. & McHeyzer-Williams, M.G. Class-switched memory B cells remodel BCRs within secondary germinal centers. *Nat. Immunol.* **16**, 296-305 (2015).

19. Pape, K.A., Taylor, J.J., Maul, R.W., Gearhart, P.J. & Jenkins, M.K. Different B Cell Populations Mediate Early and Late Memory During an Endogenous Immune Response. *Science* **331**, 1203-1207 (2011).
20. Yoshida, T. et al. Memory B and memory plasma cells. *Immunol. Rev.* **237**, 117-139 (2010).
21. Dogan, I. et al. Multiple layers of B cell memory with different effector functions. *Nat. Immunol.* **10**, 1292-1299 (2009).
22. Weisel, Florian J., Zuccarino-Catania, Griselda V., Chikina, M. & Shlomchik, Mark J. A Temporal Switch in the Germinal Center Determines Differential Output of Memory B and Plasma Cells. *Immunity* **44**, 116-130 (2016).
23. Bemark, M. et al. Limited clonal relatedness between gut IgA plasma cells and memory B cells after oral immunization. *Nature Communications* **7**, 12698 (2016).
24. Mamani-Matsuda, M. et al. The human spleen is a major reservoir for long-lived vaccinia virus–specific memory B cells. *Blood* **111**, 4653-4659 (2008).
25. Tangye, S.G. & Tarlinton, D.M. Memory B cells: Effectors of long-lived immune responses. *Eur. J. Immunol.* **39**, 2065-2075 (2009).
26. Giesecke, C. et al. Tissue Distribution and Dependence of Responsiveness of Human Antigen-Specific Memory B Cells. *J. Immunol.* **192**, 3091-3100 (2014).
27. Liu, Y.-J. et al. Memory B cells from human tonsils colonize mucosal epithelium and directly present antigen to T cells by Rapid Up-Regulation of B7-1 and B7-2. *Immunity* **2**, 239-248 (1995).
28. Dunn-Walters, D.K., Isaacson, P.G. & Spencer, J. Sequence analysis of rearranged IgVH genes from microdissected human Peyer's patch marginal zone B cells. *Immunology* **88**, 618-624 (1996).
29. Lindner, C. et al. Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota. *Nat. Immunol.* **16**, 880-888 (2015).
30. Zehentmeier, S. et al. Static and dynamic components synergize to form a stable survival niche for bone marrow plasma cells. *Eur. J. Immunol.* **44**, 2306-2317 (2014).
31. Medina, F., Segundo, C., Campos-Caro, A., González-Garćia, I.s. & Brieva, J.A. The heterogeneity shown by human plasma cells from tonsil, blood, and bone marrow reveals graded stages of increasing maturity, but local profiles of adhesion molecule expression. *Blood* **99**, 2154-2161 (2002).
32. Nguyen, D.C. et al. Factors of the bone marrow microniche that support human plasma cell survival and immunoglobulin secretion. *Nature Communications* **9**, 3698 (2018).
33. Carrion, C. et al. Adult Bone Marrow Three-Dimensional Phenotypic Landscape of B-Cell Differentiation. *Cytometry Part B: Clinical Cytometry* **96**, 30-38 (2019).
34. Apostoaei, A. & Trabalka, J.R. Review, synthesis, and application of information on the human lymphatic system to radiation dosimetry for chronic lymphocytic leukemia. *Oak Ridge, TN* **37830**, 1-64 (2010).
35. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D.R. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566**, 393-397 (2019).
36. Soto, C. et al. High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398-402 (2019).
37. Vollmers, C., Sit, R.V., Weinstein, J.A., Dekker, C.L. & Quake, S.R. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci.* **110**, 13463-13468 (2013).
38. Lavinder, J.J., Horton, A.P., Georgiou, G. & Ippolito, G.C. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Curr. Opin. Chem. Biol.* **24**, 112-120 (2015).
39. Liu, J.K.H. The history of monoclonal antibody development – Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery* **3**, 113-116 (2014).
40. Akkina, R. Humanized Mice for Studying Human Immune Responses and Generating Human Monoclonal Antibodies. *Microbiology Spectrum* **2**, 2.2.03 (2014).
41. Robinson, W.H. Sequencing the functional antibody repertoire—diagnostic and therapeutic discovery. *Nature Reviews Rheumatology* **11**, 171-182 (2015).
42. Georgiou, G. et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32**, 158-168 (2014).
43. DeKosky, B.J. et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* **31**, 166-169 (2013).
44. Busse, C.E., Czogiel, I., Braun, P., Arndt, P.F. & Wardemann, H. Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *Eur. J. Immunol.* **44**, 597-603 (2014).

45. DeKosky, B.J. et al. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat. Med.* **21**, 86-91 (2015).
46. Reddy, S.T. et al. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat. Biotechnol.* **28**, 965-969 (2010).
47. Ehling, R.A. et al. Single-cell sequencing of plasma cells from COVID-19 patients reveals highly expanded clonal lineages produce specific and neutralizing antibodies to SARS-CoV-2. *bioRxiv*, 2021.2002.2012.430940 (2021).
48. Corti, D. et al. Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. *Science* **351**, 1339-1342 (2016).
49. Jones, B.E. et al. The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates. *Sci. Transl. Med.* **13**, eabf1906 (2021).
50. Wine, Y. et al. Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. *Proc. Natl. Acad. Sci.* **110**, 2993-2998 (2013).
51. Chen, J. et al. Proteomic Analysis of Pemphigus Autoantibodies Indicates a Larger, More Diverse, and More Dynamic Repertoire than Determined by B Cell Genetics. *Cell Rep.* **18**, 237-247 (2017).
52. Lavinder, J.J. et al. Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc. Natl. Acad. Sci.* **111**, 2259-2264 (2014).
53. Lee, J. et al. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat. Med.* **22**, 1456-1464 (2016).
54. Lee, J. et al. Persistent Antibody Clonotypes Dominate the Serum Response to Influenza over Multiple Years and Repeated Vaccinations. *Cell Host Microbe* **25**, 367-376.e365 (2019).
55. Mills, J.R., Barnidge, D.R. & Murray, D.L. Detecting monoclonal immunoglobulins in human serum using mass spectrometry. *Methods* **81**, 56-65 (2015).
56. Bondt, A. et al. Human plasma IgG1 repertoires are simple, unique, and dynamic. *Cell Systems* **12**, 1131-1143.e1135 (2021).
57. Bondt, A., Dingess, K.A., Hoek, M., van Rijswijck, D.M.H. & Heck, A.J.R. A Direct MS-Based Approach to Profile Human Milk Secretory Immunoglobulin A (IgA1) Reveals Donor-Specific Clonal Repertoires With High Longitudinal Stability. *Front. Immunol.* **12**, 789748 (2021).
58. van Keulen, B.J. et al. Human Milk from Previously COVID-19-Infected Mothers: The Effect of Pasteurization on Specific Antibodies and Neutralization Capacity. *Nutrients* **13**, 1645 (2021).
59. Dingess, K.A. et al. Humans have distinct repertoires of IgA1. *bioRxiv*, 2022.2004.2008.487627 (2022).
60. van Rijswijck, D.M.H. et al. IgG1 responses following SARS-CoV-2 infection are polyclonal and highly personalized, whereby each donor and each clone displays a distinct pattern of cross-reactivity against SARS-CoV-2 variants. *bioRxiv*, 2022.2002.2024.481778 (2022).
61. Melani, R.D. et al. Next-Generation Serology by Mass Spectrometry: Readout of the SARS-CoV-2 Antibody Repertoire. *J. Proteome Res.* **21**, 274-288 (2022).
62. de Graaf, S.C., Hoek, M., Tamara, S. & Heck, A.J.R. A perspective toward mass spectrometry-based de novo sequencing of endogenous antibodies. *MAbs* **14**, 2079449 (2022).
63. Frese, C.K. et al. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **84**, 9668-9673 (2012).
64. Peng, W., Pronker, M.F. & Snijder, J. Mass Spectrometry-Based De Novo Sequencing of Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme. *J. Proteome Res.* **20**, 3559-3566 (2021).
65. Tran, N.H. et al. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **6**, 31730 (2016).
66. Sen, K.I. et al. Automated Antibody De Novo Sequencing and Its Utility in Biopharmaceutical Discovery. *J. Am. Soc. Mass. Spectrom.* **28**, 803-810 (2017).
67. Bandeira, N., Pham, V., Pevzner, P., Arnott, D. & Lill, J.R. Automated de novo protein sequencing of monoclonal antibodies. *Nat. Biotechnol.* **26**, 1336-1338 (2008).
68. Altelaar, A.F.M., Munoz, J. & Heck, A.J.R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* **14**, 35-48 (2013).
69. Guthals, A. et al. De Novo MS/MS Sequencing of Native Human Antibodies. *J. Proteome Res.* **16**, 45-54 (2017).
70. Schulte, D., Peng, W. & Snijder, J. Template-based assembly of proteomic short reads for *de novo* antibody sequencing and repertoire profiling. *bioRxiv*, 2022.2003.2007.483237 (2022).
71. Boutz, D.R. et al. Proteomic Identification of Monoclonal Antibodies from Serum. *Anal. Chem.* **86**, 4758-4766 (2014).

72. Cheung, W.C. et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat. Biotechnol.* **30**, 447-452 (2012).
73. Brodbelt, J.S., Morrison, L.J. & Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **120**, 3328-3380 (2020).
74. Fornelli, L. et al. Analysis of Intact Monoclonal Antibody IgG1 by Electron Transfer Dissociation Orbitrap FTMS. *Mol. Cell. Proteomics* **11**, 1758-1767 (2012).
75. Tsybin, Y.O. et al. Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **83**, 8919-8927 (2011).
76. Zubarev, R.A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **15**, 12-16 (2004).
77. Lermyte, F., Valkenborg, D., Loo, J.A. & Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* **37**, 750-771 (2018).
78. Cotham, V.C. & Brodbelt, J.S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. *Anal. Chem.* **88**, 4004-4013 (2016).
79. Shaw, J.B. et al. Sequencing Grade Tandem Mass Spectrometry for Top–Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Anal. Chem.* **90**, 10819-10827 (2018).
80. Mao, Y., Valeja, S.G., Rouse, J.C., Hendrickson, C.L. & Marshall, A.G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **85**, 4239-4246 (2013).
81. Shaw, J.B. et al. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **92**, 766-773 (2020).
82. Fornelli, L. et al. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **90**, 8421-8429 (2018).
83. He, L. et al. Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS. *J. Am. Soc. Mass. Spectrom.* **28**, 827-838 (2017).
84. He, L. et al. Classification of Plasma Cell Disorders by 21 Tesla Fourier Transform Ion Cyclotron Resonance Top-Down and Middle-Down MS/MS Analysis of Monoclonal Immunoglobulin Light Chains in Human Serum. *Anal. Chem.* **91**, 3263-3269 (2019).

CHAPTER 7

SELECTIVITY OVER COVERAGE IN DE NOVO SEQUENCING OF IGGS

Maurits A. den Boer[1,2,*], Jean-Francois Greisch[1,2,*], Sem Tamara[1,2,*], Albert Bondt[1,2], and Albert J.R. Heck[1,2]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

[2] *Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands*

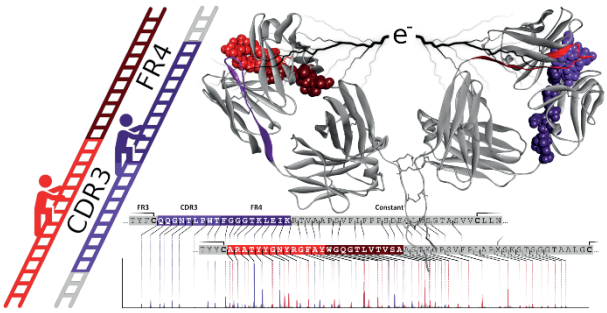
[*] *These authors contributed equally*

Published as:

Selectivity over Coverage in De Novo Sequencing of IgGs
Maurits A. den Boer, Jean-Francois Greisch, Sem Tamara, Albert Bondt, and Albert J.R. Heck
Chemical Science 2020 11, 11886
DOI: 10.1039/d0sc03438j

Abstract

Although incredibly diverse in specificity, millions of unique Immunoglobulin G (IgG) molecules in the human antibody repertoire share most of their amino acid sequence. These constant parts of IgG do not yield any useful information in attempts to sequence antibodies *de novo*. Therefore, methods focusing solely on the variable regions and providing unambiguous sequence reads are strongly advantageous. We report a mass spectrometry-based method that uses electron capture dissociation (ECD) to provide straightforward-to-read sequence ladders for the variable parts of both the light and heavy chains, with a preference for the functionally important CDR3. We optimize this method on the therapeutic antibody Trastuzumab and demonstrate its applicability on two monoclonal quartets of the four IgG subclasses, IgG1, IgG2, IgG3 and IgG4. The method is based on proteolytically separating the variable $F(ab')_2$ part from the conserved Fc part, whereafter the $F(ab')_2$ portions are mass-analyzed and fragmented by ECD. Pure ECD, without additional collisional activation, leads to straightforward-to-read sequence tags covering the CDR3 of both the light and heavy chains. Using molecular modelling and structural analysis, we discuss and explain this selective fragmentation behavior and describe how structural features of the different IgG subclasses lead to distinct fragmentation patterns. Overall, we foresee that pure ECD on $F(ab')_2$ or Fab molecules can become a valuable tool for the *de novo* sequencing of serum antibodies.



Introduction

Antibodies are key molecules of our immune system and therefore of great interest to many researchers, both in academia and industry[1-4]. The high selectivity of immunoglobulins G (IgGs) for a single antigen has made this class of antibodies of great value in basic research, but even more important in molecular medicine[5]. Therapeutic antibodies reflecting those of an individual having survived a severe infection, *e.g.*, by the SARS-CoV-2, Dengue[6] or Ebola virus[7], can lead to a substantial reduction in the mortality rate of newly infected patients. Shortening the route from the discovery of an antibody to the production of a functional equivalent is therefore of utmost importance, certainly in present times[8].

Structurally, IgGs are made of four disulfide-bridged polypeptide chains: two identical light chains (LC) of 25 kDa and two identical heavy chains (HC) of approximately 50 kDa (**Figure 1A**). The LC constant domain can be of two types, namely kappa (κ) or lambda (λ). The HC constant domain, on the other hand, is divided into four subclasses, namely IgG1, IgG2, IgG3, and IgG4, having between 83% and 96% sequence similarity. Only specific stretches of the HC regions differ between the subclasses: mainly the hinge region and the N-terminal side of the HC region below the hinge (Fc). Additionally, IgG subclasses exhibit distinctive disulfide bridging with a different number of disulfide bonds in the hinge region and different positioning of disulfides between LC and HC[9,10].

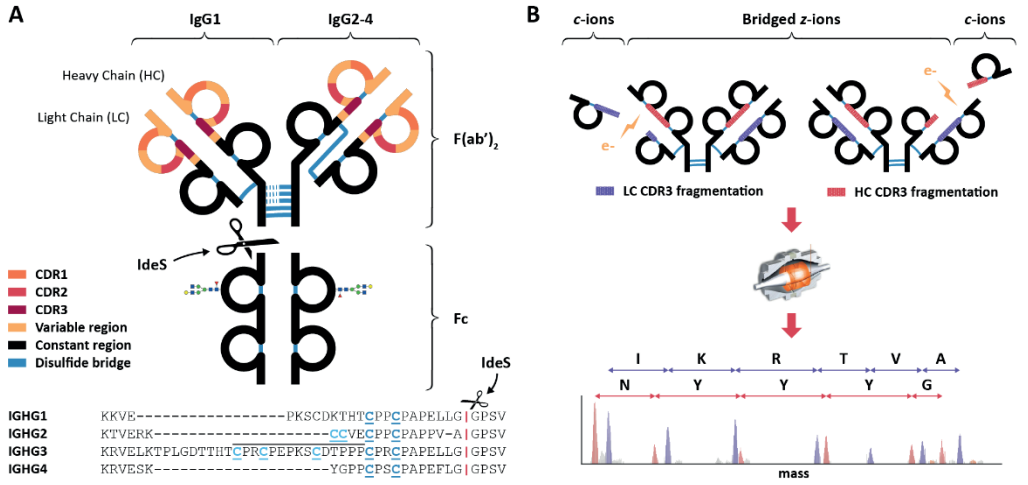


Figure 1 | Schematic overview of IgGs structures and the here used sequencing workflow. (A) Distinctive structures of IgG1 and IgG2-4, all consisting of two variable regions (colored) with three antigen-binding Complementarity Determining Regions (CDRs) each, and one constant region (black). The first step in the workflow involves the separation by proteolytic cleavage of the IgG CDR-containing F(ab')$_2$ from the glycosylated constant Fc using the IdeS enzyme, which cleaves (scissors) just below the disulfides bridges (blue) between the heavy chains (HCs). Besides different numbers of HC-bridging disulfides, IgG1-4 subclasses also differ by the disulfide bond between the light chain (LC) and the HC: directly above the hinge for IgG1, mediated by a more N-terminal HC cysteine for IgG 2-4. Additionally, the IgG3 hinge region is longer owing to an insertion that is repeated three times, as indicated with a black line above the sequence. **(B)** The purified F(ab')$_2$ portions of all IgG subclasses are subjected to ECD-MS, which preferentially fragments the LC and HC between their disulfide-bridged loops. Clearly separated low *m/z* *c*-ions and complementary high *m/z* bridged *z*-ions are formed, enabling the straightforward interpretation of fragment ion ladders beneficial for *de novo* IgG sequencing.

Antigen recognition by IgG is mainly determined by three hypervariable regions located in the N-terminal domains of both the LC and HC. Of these so-called complementarity determining regions (CDR1, CDR2 and CDR3), the two CDR3s are often assumed to be major determinants of antibody-binding specificity[11,12]. Genetic recombination and hypermutation in these CDR regions lead to a plethora of possible human IgG antibodies, whereby it has been postulated that humans can generate about 10 billion unique antibodies, each potentially capable of binding an antigenic determinant (epitope) with high selectivity[13].

This amazingly large repertoire is possible because the CDR sequences are only partly genome derived. After recombination of multiple genome segments into one antibody sequence, so-called V(D)J-recombination, random mutations are introduced in the process of somatic hypermutation. While initially used to increase the spectrum of recognized antigens, these processes subsequently lead to affinity and selectivity enhancements.

A common approach to elucidate the antibody repertoire in a human is to sequence the immunoglobulin mRNAs from memory B cells present in the donor's circulation[14]. The majority of antibodies present in the circulation is however not produced by these circulating B cells, but by long-lived plasma cells that reside in the bone marrow, which are much harder to harvest. The most accessible sequence-containing unit of relevant circulating antibodies is, therefore, the protein itself[15].

Sequencing proteins is nowadays generally performed by mass spectrometry. Characterization of peptides of 5 to 20 amino acids produced by proteolytic cleavage results into correct annotations of the full polypeptide backbone, even for *de novo* approaches when no DNA or protein sequence is available[16-20]. However, when it comes to antibodies from mixtures such as present in serum, sample complexity rapidly increases, preventing high-confidence assignments[15]. One way to overcome this hurdle is to apply top-down proteomics since the hypervariable region of interest – the paired N-terminal domains of the LC and HC – is then left intact and can be isolated[21]. The relevant information density increases even further when the constant regions, or at least most of the constant parts containing PTMs such as lysine clipping or glycosylation, are removed.

Here, we hypothesized that the ideal approach for *de novo* sequencing of antibodies from complex mixtures is a top-down or middle-down mass spectrometry-based method that results in long stretches of straightforward-to-read sequence ladders covering as much of the variable domains as possible, but at least the CDR3. This would resemble a modern, mass spectrometry-based, version of Edman degradation capable of dealing with impurities or complex mixtures and allowing for longer sequence reads. The work presented here is aimed at such a methodology.

Straightforward-to-read sequence ladders require highly optimized mass spectrometric methods preferentially yielding only one type of fragment pairs (a/x, b/y, or c/z) to avoid spectral congestion and dilution of signal intensity. To this end, electron capture dissociation (ECD)[22], which recently became available for benchtop mass spectrometers, has been shown to offer significant advantages in obtaining straightforward-to-read sequence ladders[23,24]. Since, ECD has been used to query antibody sequences. It resulted

in significant backbone cleavages in both the LC and HC variable regions of both denatured and native intact antibodies. It also yielded mainly ($c/z\cdot$) fragment ions[25,26]. While the presence of intrachain disulfide bridges complicates the detection of several variable domain regions, it enhances the detection of CDR3 fragments[25,27].

In the present work, we build upon these exciting results and further optimize ECD for the *de novo* sequencing of IgG CDR3s and their downstream regions, generating very clean, straightforward-to-read sequence ladders comprised solely of *c*-ions (**Figure 1B**). Using native conditions in electrospray to ionize the IgG antibodies, we achieve maximal separation between the precursor and the informative fragment ions. We show that removal of the Fc region also reduces the complexity of the spectrum and simplifies the precursor ion selection. We optimize and apply the method using the IgG1 Trastuzumab. Next, in order to investigate in detail whether the method is generically applicable, we apply it to two quartets of IgG covering all human IgG subclasses (*i.e.*, IgG1, IgG2, IgG3 and IgG4). We show that straightforward sequence reads are obtained for all IgG subclasses. Through structural modeling, we aim to explain and predict the observed fragmentation patterns.

Experimental

Antibody samples

Monoclonal antibodies against CD20 (7D8)[28] and DNP (DNP-G2a2)[29,30] were recombinantly expressed as IgG1, IgG2, IgG3 and IgG4, and provided by Genmab (Utrecht, the Netherlands). The IgG1 antibody Trastuzumab was provided by Roche (Penzberg, Germany).

Preparation of intact IgG, F(ab')$_2$ and Fab samples for native top-down MS analysis

F(ab')$_2$ molecules were generated by overnight digestion at 37 °C of the IgGs bound to a CaptureSelect FcXL affinity matrix (Thermo Fisher Scientific) using 0.25 U/µg FabRICATOR® IdeS (Genovis AB, Llund, Sweden) as described earlier[31]. IgG1 Fab molecules were similarly prepared using 1 U/µg FabALACTICA® IgdE (Genovis AB, Llund, Sweden). Prior to native top-down analysis, buffers were exchanged to aqueous 150 mM ammonium acetate (pH 7.5) through six consecutive dilution and concentration steps at 4 °C using Amicon® Ultra centrifugal filters with a 10 kDa molecular weight cutoff (Merck KGaA, Darmstadt, Germany). IgG concentrations used for native electrospray ionization mass spectrometry were around 4 µM.

Native top-down ECD MS

Top-down MS of native IgG (sub)complexes was performed on an ultra-high mass range (UHMR) Q-Exactive Orbitrap (Thermo Fisher Scientific, Bremen, Germany) using an ECD cell developed by e-MSion[32]. First, IgG, F(ab')$_2$ or Fab precursor ions were mass-selected by the quadrupole. The in-source trapping desolvation voltage displays an optimum in terms of ECD at -100 V, likely a consequence of small structural changes and cleavage enhancement occurring as the effective temperature of the ion is raised[33]. The ions were then transmitted into the ECD cell where they were subjected to electron capture

dissociation. The small permanent magnets ensure the radial confinement of electrons emitted by a cathode while additional electrodes ensure longitudinal confinement and efficient ion transfer to the HCD cell. Second, upon transfer from the ECD cell to the HCD cell, post-ECD collisional activation was kept to a minimum (HCD direct eV setting = 1) to avoid (*b*, *y*) ion formation and concentrate the fragment signal intensity into the *c*- and *z*-ions of interest. Overall, the HCD cell trapping and extraction parameters were optimized for low nitrogen collision gas pressures to achieve efficient detection of the subunits and fragments. All spectra were acquired with the noise threshold parameter set to 3.64. A single charge state of the precursor (21+ for F(ab')$_2$, 13+ for F(ab), and 24+ for intact IgG) was isolated to simplify interpretation of the fragment ion spectra. All the spectra were acquired at a set resolution of 200,000 @ *m/z* 400.

Data analysis

Processing of the fragmentation spectra involved the conversion of raw files to mzML format by Proteowizard[34]. We used the MSDeisotope python library (Joshua Klein, Boston University CBMS)[35,36] with a minimum_score = 10.0 and mass_error_tolerance = 0.02 to generate a charge deconvoluted spectrum with all the isotopic peaks retained[35,36]. Fragments were assigned by applying LcMsSpectator (Pacific Northwest National Laboratory)[37,38] to the charge deconvoluted spectra generated by MSDeisotope. The accuracy threshold was set to 3 ppm for all assignments following recalibration of the fragment's *m/z* via shifting by the average error on assignable *c*-ion fragments. The results were exported as tsv files for further analysis. Sequence assignment accommodated the major ECD ion types (*c*, *z*, *z*·) without considering H_2O and NH_3 neutral losses, except when explicitly mentioned.

Modelling

The structures of the F(ab')$_2$ fragments studied are not available in structural databases. Consequently, homology models of the F(ab')$_2$ were built using MODELLER[39,40] similarly to the protocol introduced by Politis and co-workers[41]. For that, we used crystallographic structures with PDB IDs IgG1 1HZH (human)[42], IgG2 1IGT (mouse)[43], IgG3 5A16 (mouse)[44] and IgG4 5DK3 (human)[45] as a template. Refinement of the obtained models was also performed in MODELLER using a combination of variable target function method and molecular dynamics with simulated annealing; optimization was repeated for three cycles. All models were assessed using the (negative) Discrete Optimized Protein Energy score[39,40]. Since IgG3 lacks an available F(ab')$_2$ structure, the hinge region was built in several steps to avoid knots introduced by MODELLER when more than approximately 15 residues are missing from the template structure. All disulfide bridges were checked and manually defined when they were not modelled automatically.

To explore the conformational space of the Fab arm movement, each F(ab')$_2$ model was subjected to a rapidly exploring random tree based sampling, which is part of the Integrative Modelling Platform[46]. For that, residues in the upper hinge region were set as flexible linker residues, namely: D226-T230 for anti-CD20 IgG1 and D225-T229 for anti-DNP IgG1, E221-K223 for anti-CD20 IgG2 and E220-K222 for anti-DNP IgG2, E221-T232 for anti-CD20 IgG3 and E220-T231 for anti-DNP IgG3, E221-P227 for anti-CD20 IgG4 and E220-P226 for anti-DNP IgG4. The rest of the molecule was connected through inter-chain disulfides and set as a rigid body. Next, the algorithm randomly tested the possibility of

rotating an atom of the linker residues in agreement with the torsional space of the corresponding residue. The sampling was performed with the number of iterations and nodes set to 10,000. Subsequently, conformations without steric clashes and overlaps were used as output. High-throughput collision-cross section (CCS) calculations were done using Impact software[47] using the standard parameters. The obtained CCS values were used for selecting the appropriate set of F(ab')$_2$ conformations. Calculation of the center-of-mass (COM) of the light chain, heavy chain and the F(ab')$_2$ was performed in PyMOL[48] on a sample of 2500 random and 500 most compact conformations. For COM calculations of the IgG3 F(ab')$_2$, we used HC sequences down to the first two hinge disulfides to avoid the COM offset introduced by the larger hinge region of IgG3, when compared to other IgG subclasses. Next, we calculated the average distances between the COM of the LC or HC and the COM of the F(ab')$_2$ for a subset of conformations to assess the degree to which each chain is either inward- or outward-facing. Finally, we used this metric to estimate which chain is likely to be more exposed to the environment when the collapsed gas-phase structures are formed.

Results and Discussion

Benchmarking ECD-only on intact IgG1, F(ab')$_2$, and Fab molecules of Trastuzumab

At the start of our experiments, we tested the performance of ECD on different preparations of Trastuzumab. A cross-comparison between ECD fragment ion spectra obtained for the 148 kDa intact IgG1, the 98 kDa F(ab')$_2$ and the 48 kDa Fab molecules revealed a striking similarity (**Figure S1** and **Table S1**). In line with previously reported data[27], we observed that most of the abundant fragment ions have masses between 9000 and 16,000 Da, regardless of the Trastuzumab moiety that was used.

After careful consideration and examination of the ECD fragmentation spectra, we chose to focus on F(ab')$_2$ molecules for the rest of our study for the reasons described below. The ECD fragment ion spectra of intact Trastuzumab were notably more complex to analyze due to spectral congestion by *z*-ions from the highly conserved Fc region, which overlap with the *c*-ions from the region of interest. Additionally, the glycosylation on the Fc dilutes the precursor ion signal over a broader range of *m/z* values, hampering the use of narrow *m/z* windows for precursor isolation. Therefore, we chose to disfavor the analysis of intact mAbs.

Although both Fab and F(ab')$_2$ preparations yield straightforward sequence ladders corresponding to *c*-ions from the LC and HC regions of interest, the F(ab')$_2$ molecules provide some key advantages. While the IgdE enzyme can reliably produce Fab molecules of IgG1[49], other IgG subclasses require the use of less specific enzymes such as papain, which result in higher sample heterogeneity. The IdeS enzyme on the other hand, which generates F(ab')$_2$ molecules by cleaving at a specific position below the hinge of all IgG subclasses[50], provides both a reproducible and universal method irrespective of the subclass investigated. The resulting low-heterogeneity samples derived for the F(ab')$_2$ preparations enable strict precursor isolation without impacting sensitivity. Additionally, for denatured samples as well as native Fabs, the relatively low *m/z* of the precursor ions

place them in the same *m/z* window where the ECD fragments are observed. The use of native F(ab')$_2$ molecules concentrates the precursor ion intensity on a small number of charge states outside the fragment ion regions, thereby simplifying data analysis. For these reasons, we chose the native F(ab')$_2$ molecules as precursors in the remainder of our analysis, although our data confirm that ECD also works reasonably well for both intact mAbs as well as Fab fragments.

Supplemental activation reduces the selectivity of ECD

Combining ECD with supplemental collisional activation can enhance precursor fragmentation. Although this has been demonstrated by Shaw *et al.*[27] to extend sequence coverage, collisional activation also generates additional *b*/*y*-ion series. We argue that the generation of these extra fragment ion series could be disadvantageous for straightforward and confident *de novo* sequencing of IgGs.

We investigated the effects of supplemental activation by recording EChcD-MS spectra of F(ab')$_2$ molecules using different levels of collisional activation. The higher collision voltages used in EChcD indeed introduced other ion types, which increase spectral complexity and make it more difficult to extract unambiguous sequence reads (**Figure S2** and **Table S2**). Moreover, the slightly improved sequence coverage obtained by EChcD primarily provided information on the highly conserved parts of the antibody sequence, notably the FR3 and constant regions.

Since *de novo* sequencing benefits from non-ambiguous sequence reads, we opted for the use of "pure" ECD (*i.e.*, without HCD) to reduce the fragmentation pattern to the single *c*-ion type covering the regions of interest. Notably, after careful analysis of all fragment ions with masses in between 9000 and 16,000 Da, we could conclude that almost all fragment ion intensity originated exclusively from *c*-ions, with the low-abundant remainder originating from *a*-ions. Thus, conditions can be employed wherein ECD of F(ab')$_2$ molecules purely generates *c*-ions in this mass window.

Pure ECD produces clean CDR3-spanning *c*-ion ladders

Following our encouraging benchmark data obtained for the F(ab')$_2$ molecules of Trastuzumab, we sought to determine whether the observed fragmentation patterns were generic and thus useful for a range of different IgG mAbs. To this end, we characterized two quartets of IgG mAbs, covering the subclasses IgG1, IgG2, IgG3 and IgG4. One of these mAbs targets CD20, the other 2,4-dinitrophenol (DNP). Sequences, as well as theoretical and observed masses, are provided at the end of the Supporting Information in **Table S5** and **Table S6**.

Application of pure ECD to F(ab')$_2$ molecules of all subclasses of the two IgG families consistently yielded ECD fragment ion mass spectra containing pure and easily interpretable *c*-ion sequence ladders covering CDR3 and FR4 of both the LC and HC (**Figure 2**, **Figure S3**, and **Table S3**). We noticed that the fragment ions originate nearly exclusively from backbone cleavages in a part of the LC and HC sequence that is framed by

intramolecular disulfide loops, and which span from the beginning of the CDR3 to the constant region. For both LC and HC, the observed backbone cleavages start immediately after the cysteine residue preceding the CDR3 (LC: Cys-88, HC: Cys-96). This resulted in near-complete sequence coverage of this very important region, with the exception that we observed no *c*-ion cleavages at the N-terminal side of proline residues, in line with earlier reports[51-53]. Notably, several of these gaps are covered by the corresponding *a*-ions, although at relatively lower intensities than the *c*-ion series.

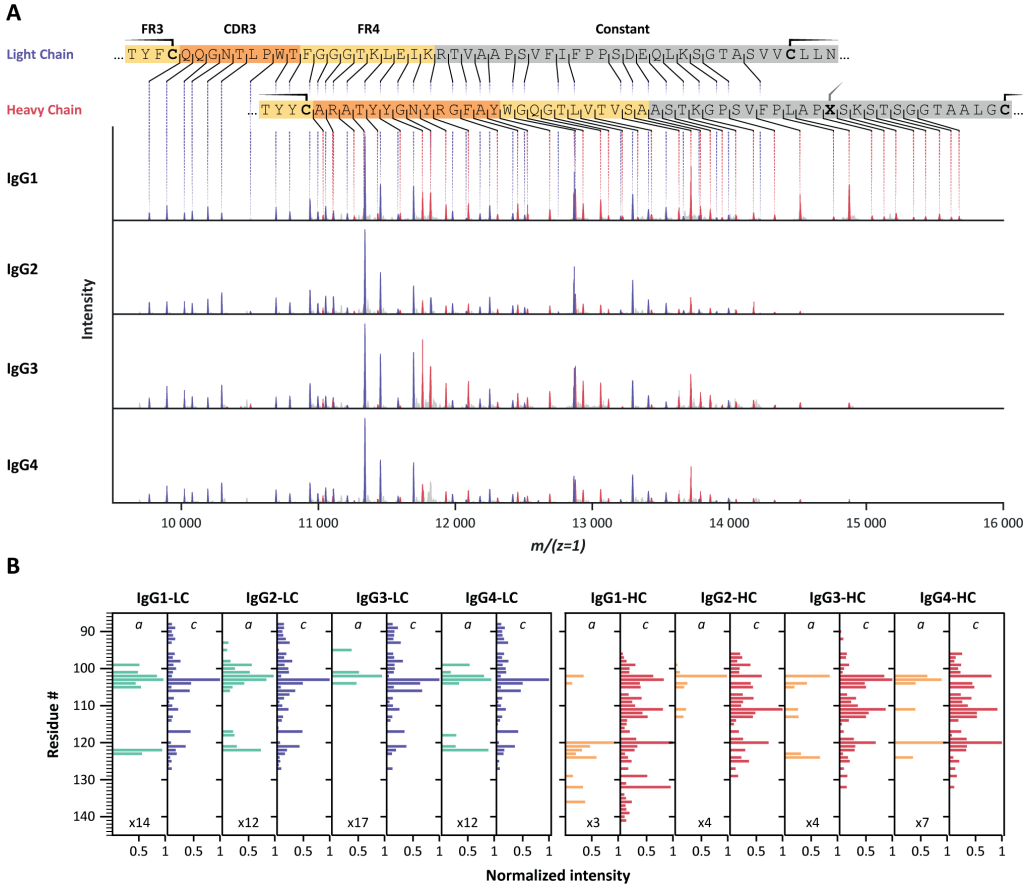


Figure 2 | Pure ECD of anti-DNP IgG F(ab')$_2$ molecules produces clean CDR3-spanning *c*-ion ladders. **(A)** Charge-deconvoluted ECD-MS spectra of IgG1, IgG2, IgG3 and IgG4 F(ab')$_2$ molecules of anti-DNP showing exclusively *c*-ion ladders covering the LC (purple) and HC (red) CDR3 regions. Cysteine residues involved in disulfide bonds are highlighted on the color-coded sequences by a solid line. Residue X, in …PLAPXS… in the heavy chain, represents a Ser for IgG1, and a Cys for IgG2-4. **(B)** Bar chart displaying normalized *c*- and *a*-ion fragment intensities for each residue of IgG1, IgG2, IgG3 and IgG4 LC and HC. Although all ECD spectra for a given mAb are very similar, IgG1 HC *c*-ion fragments further extend towards the C-terminus for both investigated mAb families (Figure S3). For some cleavage sites, *a*-ions complement missing *c*-ions.

Ion ladders covering the CDR3 of the LC do not overlap with HC fragment ions in the charge-deconvoluted spectra, making them suitable for direct interpretation. Ion ladders covering the CDR3 of the HC, on the other hand, overlap with fragment ions covering FR4 and the constant region of the LC. Because these regions are highly conserved, a comparison with theoretical spectra from these regions, accounting for the unknown N-terminal part by a variable mass shift, could help disentangle LC and HC fragment ions, making these mass spectra very suitable for *de novo* sequence assignment.

Exploring fragment complementarity in pure ECD of IgG F(ab')$_2$ molecules

By using low pressure in the HCD cell and avoiding collisional activation, the ECD fragments of the two IgG quartets nearly exclusively correspond to *c*-ions (**Table S4**), similar to our data for Trastuzumab, without the need to invoke small neutral losses such as ammonia or water, or sidechain losses. Apart from *c*-ions, a minor contribution is made by *a*-ions, often when there is no *c*-ion fragment.

Although C-terminal *z*-ions are not apparent within the shown deconvoluted analysis window (9000 < Mw < 16,000 Da) of **Figure 2**, they were detected completely separated in *m/z* from the concomitant *c*-ions (**Figure 3A**). This is because the formed *z*-ions remain in complex with the remaining F(ab')$_2$ precursor, as they are connected through disulfide bridges at or near their C-termini. Charge and mass determination support this conclusion. While the 9-16 kDa *c*-ion fragments carry approximately 3-8 charges, their concomitant 83-90 kDa bridged *z*-ion fragments carry approximately 8-17 charges (**Figure 3B**). Overall, pure ECD of F(ab')$_2$ molecules results specifically in a small subset of low-mass *c*-ions and corresponding high-mass bridged *z*-ions.

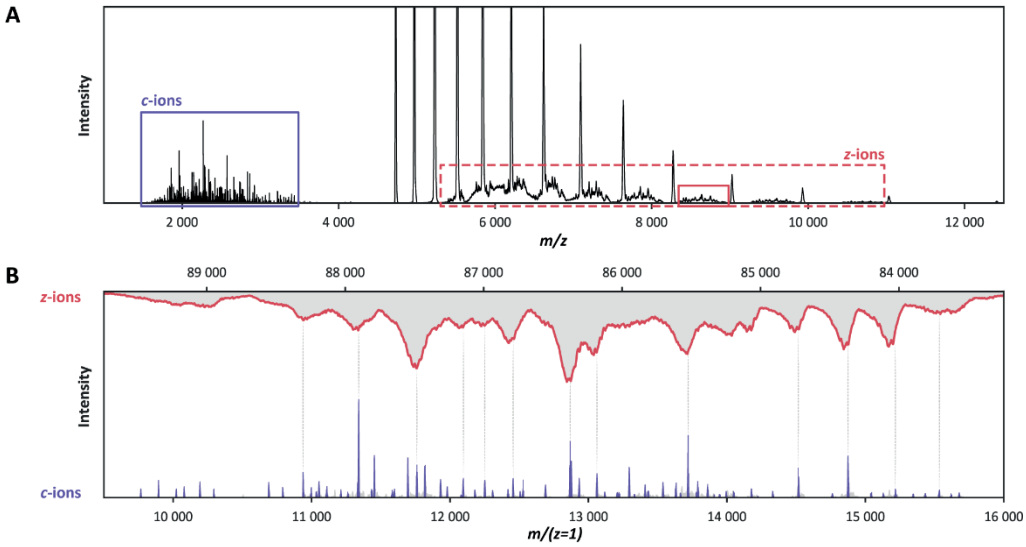


Figure 3 | Concomitant N-terminal *c*-ions and C-terminal bridged *z*-ions produced in the ECD spectra of the anti-DNP IgG1 F(ab')$_2$ precursor, revealing that the disulfide bonds remain intact. (A) The *c*-ions (purple box) are well separated in the *m/z* window from their concomitant bridged *z*-ions (red dotted box) in the native ECD-MS spectrum of the IgG1 anti-DNP F(ab')$_2$ molecule. Because they are minimally affected by signal overlapping, bridged *z*-ions in the solid red box were selected for further analysis. **(B)** Alignment of the *c*-ions (purple) with the bridged *z*-ions (red) after charge deconvolution reveals that their mass adds up to that of the intact 99.2 kDa F(ab')$_2$ precursor.

The disulfide links present in the IgG subclasses direct specific ECD fragmentation

The intact IgGs, but also the F(ab')$_2$ and Fab molecules, contain hundreds of amino acid residues, which could all theoretically be cleaved by ECD, potentially leading to hundreds of *c*- and *z*-ions. Still, we observe that, very selectively and reproducibly, only a few dozen specific fragment ions are preferentially formed. To understand this distinct fragmentation pattern, observed in all ECD spectra of F(ab')$_2$ molecules from the different IgG subclasses, we next looked into their (distinct) structural features. Since exact structural models of the studied F(ab')$_2$ molecules are not available, we performed homology modelling, whereby available IgG structures were used as templates for the structural reconstruction of the target F(ab')$_2$ molecules. The high quality of the resulting F(ab')$_2$ structural models, as displayed in **Figures 4** and **S5,** can be assessed from their Discrete Optimized Protein Energy (DOPE) profiles (**Figure S4**). An overview of the tools used for the structural analysis of the here studied IgG F(ab')$_2$ molecules is provided in **Figure S6**.

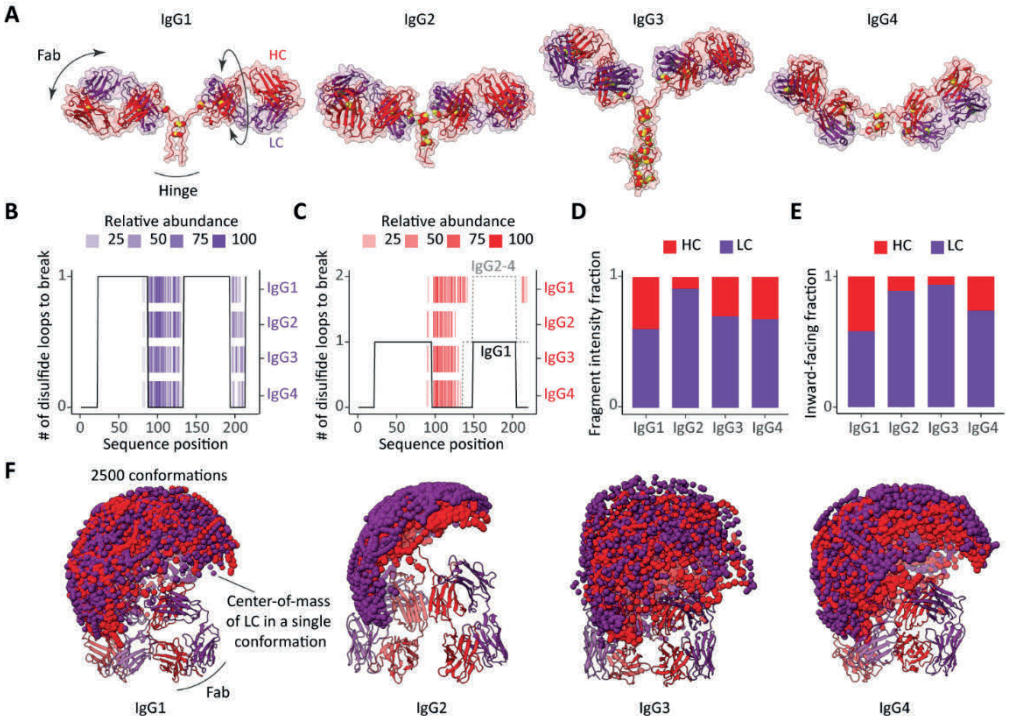


Figure 4 | Influence of structural features on ECD fragmentation of F(ab')$_2$ from the four subclasses of IgGs (anti-CD20). (A) Structures of F(ab')$_2$ comparatively modelled against available structures of IgG subclasses, IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3. The disulfide bridges are displayed as paired yellow spheres. **(B, C)** The number of disulfide bonds that span each sequence position of (**B**) the LC and (**C**) the HC in different subclasses of IgG anti-CD20 overlaid with detected ECD fragments. **(D)** Fractions of ECD fragment ion intensities originating from the LC or HC for the different subclasses of IgG anti-CD20. **(E)** Fractions of the 500 most compact conformations of each subclass in which either the LC or HC is most solvent-exposed following collapse[54]. See **Figure S7** for a more detailed description of the estimated chain relative exposure based on center-of-mass calculations. **(F)** Sampling of Fab arm movements displayed as spheres representing the center-of-mass of (purple) LCs and (red) HCs, showing remarkable differences in between the different subclasses.

The disulfide bridges are among the most distinctive structural features that differentiate one IgG subclass from another, and not surprisingly these linkages affect the observed ECD cleavages in intact IgG and F(ab')$_2$ molecules most notably[25]. In **Figure 4A** (anti-CD20) and **Figure S5A** (anti-DNP), the distinctive morphologies of the IgG subclasses are readily distinguishable. Differences are primarily in the distribution of the disulfide bridges and the length of the hinge regions, a notable distinctive feature between the IgG subclasses.

To assess the impact of the disulfide linkages on the ECD, we first assessed the number of disulfide bond cleavages required to produce N-terminal fragments of either the LC or HC. Overlaid with the *c*-ion patterns produced by ECD of F(ab')$_2$ molecules, the resulting disulfide profiles nicely explain nearly all detected ECD fragments for both the LCs and HCs (**Figure 4B**, **C** for anti-CD20 IgGs and **Figure S5B**, **C** for anti-DNP IgGs). In other words,

only backbone cleavages in the regions that are not part of an intra- or interchain disulfide loop result in separation and detection of free N-terminal ions.

Since the LCs of all subclasses within a given IgG quartet are identical, it is not surprising that they yield identical ECD fragments with comparable abundances (**Figure 2** and **Figure S5B**). The ECD patterns of HCs differ between the IgG subclasses (**Figure 2** and **Figure S5C**), but this is readily explained by the location of the LC-HC interchain disulfide bridges. In the HC sequences of IgG2-4, the interchain link is mediated by Cys-136, which is in the middle part of the sequence. For the IgG1 HC on the other hand, the interchain link is mediated by the more C-terminal Cys-225. As a result, a larger part of the IgG1 HC sequence is free from disulfide looping, and therefore susceptible to ECD.

Probing distinctive structural dynamics of F(ab')$_2$ originating from different IgG subclasses

In addition to the striking disulfide-directed fragmentation, we noticed that the total ion intensity of HC fragments is higher for IgG1 compared to other subclasses, with IgG2 displaying the lowest fraction of HC fragment ion intensity (**Figure 4D** and **Figure S5D**). We hypothesized that a possible explanation could involve differences in the conformations and structural dynamics of the IgG molecules[33,41,54,55].

In the absence of extensive collisional activation, low energy (< 1 eV) electron capture dissociation is known to result in the preferential release of fragments from the surface-exposed regions of native proteins[17,45–47]. The rationale is that exposed domains and chains have fewer non-covalent interactions upon transition to the gas phase, and therefore more readily release backbone fragments upon electron attachment. Thus, we set out to probe the structural features of F(ab')$_2$ molecules using modelling.

It has recently been reported that IgG molecules undergo a substantial structural collapse upon transfer to the gas phase during electrospray ionization[41,54]. In this structural rearrangement, the high flexibility of the hinge region enables one or more Fab arms to clasp the IgG Fc region[41]. When processed by the IdeS enzyme, the remaining F(ab')$_2$ portion of IgG retains its hinge region and, hence, F(ab')$_2$ should also be able to adopt a similar more compact structure upon transition to the gas phase. Although we are not able to directly predict the exact collapsed state of IgG F(ab')$_2$ molecules, we can assess their conformational space prior to the collapse for which we used rapidly exploring random tree based conformational sampling from the Integrative Modelling Platform[46]. First, we sampled the motions of the Fab arms (n = 2500, **Figure 4F** and **Figure S5F**). The differences in conformational dynamics between the different IgG subclasses can be assessed by displaying the center-of-mass (COM) of both chains of a Fab arm for a conformation sample[55]. In order to infer the relative position of LC and HC in the putative F(ab')$_2$ gas-phase conformations, we calculated the average distances between the COM of each chain and the COM of the molecule (**Figure S7**). Second, we considered the 500 most compact F(ab')$_2$ conformations as determined from their collision cross-sections. We then assumed that the most exposed chains, as inferred from the COMs, are less prone to end up buried within the collapsed structures and will, therefore, be more likely to capture

electrons and release backbone fragments. Finally, we compared all subclasses for anti-CD20 IgG (**Figures 4E** and **4F**) and anti-DNP IgG (**Figures S5E** and **S5F**).

Our modelling data suggest that the LCs and HCs are nearly equally solvent-exposed in the collapsed structures of IgG1 F(ab')$_2$, resulting in comparable abundances of the LC and HC ECD fragments. The IgG4 F(ab')$_2$ exhibits slightly less exposed HCs which seemingly results in a higher fraction of LC fragments. Notably, the relative rigidity[55] and predominant presence of LCs on the surface of collapsed IgG2 F(ab')$_2$ structures correlate pleasingly well with the prevalence of LC fragments over HC fragments, which explains the lower abundance of HC fragments compared to the other subclasses. Owing to its larger and more flexible hinge region, F(ab')$_2$ of IgG3 shows the widest range of Fab arm motions and the broadest distribution of collision cross-sections (**Figure S8**). It is likely that the high flexibility of IgG3 F(ab')$_2$ leads to dramatic structural rearrangements upon transfer to the gas phase, which can affect fragment separation upon ECD. Summarizing the modelling data, in IgG2, IgG3 and IgG4, the LC is more exposed than the HC while, in contrast, IgG1 displays nearly equal surface occupation for the LC and HC when sampling the 500 most compact F(ab')$_2$ conformations. Notably, our analysis supports that the predominant presence of the LC on the surface of the collapsed IgG2 F(ab')$_2$ structures and the relative rigidity of the overall structure hamper the formation of abundant HC fragments compared to the other subclasses of both anti-CD20 and anti-DNP IgG.

Conclusions

Here, we explored the benefits of electron capture dissociation in attempting to sequence antibodies. Differentiating our approach from earlier attempts[21,27,56], we aimed deliberately for a very restricted, albeit targeted, sequence coverage, spanning just the important hypervariable CDR3 of both the light and heavy chain (LC and HC) of the mAbs. We optimized our method by first getting rid of the sequence-wise less informative constant regions (Fc) by using the IdeS protease, to obtain F(ab')$_2$ fragments. Next, we conducted ECD at minimal collisional activation and low pressure, allowing us to get exclusively straightforward-to-read amino acid sequence ladders of *c*-ions spanning the CDR3s. The ladders we obtained provide easily interpretable sequence reads covering the CDR3 of both chains, whereby we seemingly only observe a gap in the sequence when there is a proline present. We argue that this optimized method can be very powerful and beneficial for the *de novo* sequencing of human antibodies.

Next, we sought to explain the observed highly specific fragmentation behavior by studying whether it was generic. Therefore, we investigated two quartets of mAbs (one directed against CD20 and one against DNP) being expressed in all their subclass variants; IgG1, IgG2, IgG3 and IgG4. On the one hand, the ECD data of these eight mAbs showed extremely conserved fragmentation of their CDR3s with the exclusive formation of *c*-ions in the low mass region (9000 - 16,000 Da). On the other hand, specific features were observed between the different subclasses, which could be largely explained by differences in the disulfide bridges within these molecules, and by their distinctive structural features in the context of the gas-phase transition.

Through this work, we improve our understanding of the observed very specific fragmentation in the CDR3 regions of all different subclasses of the IgGs. The models generated based on this data provide a first step in predicting the fragmentation patterns observed in ECD of IgGs. Such a spectrum prediction tool would greatly facilitate *de novo* sequencing of antibodies.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the members of the Heck laboratory for general support, especially Arjan Barendregt. JFG thanks Joshua Klein (Boston University CBMS) for modifications of the MSDeisotope python library. This research received funding through the Netherlands Organization for Scientific Research (NWO) TTW project 15575 (Structural analysis and position-resolved imaging of macromolecular structures using novel mass spectrometry-based approaches), the ENPPS.LIFT.019.001 project (AJRH and JFG), the NACTAR project 16442 (AJRH and MAdB) and the Spinoza Award SPI.2017.028 to AJRH. This project received additional funding from the European Union's Horizon 2020 research and innovation program under grant agreement 686547 (EPIC-XS) for AJRH. We kindly acknowledge the team of Janine Schuurman and Frank Beurskens of Genmab, Utrecht, NL for continuous support over the years and the gift of the two quartets of IgG subclass antibodies, and Dietmar Reusch and Markus Haberger from Roche, Penzberg, Germany for the gift of Trastuzumab.

References

1. Schwab, I. & Nimmerjahn, F. Intravenous immunoglobulin therapy: how does IgG modulate the immune system? *Nature Reviews Immunology* **13**, 176-189 (2013).
2. Schroeder, H.W. & Cavacini, L. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology* **125**, S41-S52 (2010).
3. Wagner, E. et al. Determination of size variants by CE-SDS for approved therapeutic antibodies: Key implications of subclasses and light chain specificities. *Journal of Pharmaceutical and Biomedical Analysis* **184**, 113166 (2020).
4. Giorgetti, J., Beck, A., Leize-Wagner, E. & François, Y.-N. Combination of intact, middle-up and bottom-up levels to characterize 7 therapeutic monoclonal antibodies by capillary electrophoresis – Mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **182**, 113107 (2020).
5. Leavy, O. Therapeutic antibodies: past, present and future. *Nature Reviews Immunology* **10**, 297-297 (2010).
6. Dejnirattisai, W. et al. A new class of highly potent, broadly neutralizing antibodies isolated from viremic patients infected with dengue virus. *Nat. Immunol.* **16**, 170-177 (2015).
7. Mulangu, S. et al. A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *New England Journal of Medicine* **381**, 2293-2303 (2019).
8. Iwasaki, A. & Yang, Y. The potential danger of suboptimal antibody responses in COVID-19. *Nature Reviews Immunology* **20**, 339-341 (2020).

9. Vidarsson, G., Dekkers, G. & Rispens, T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Front. Immunol.* **5** (2014).
10. Liu, H. & May, K. Disulfide bond structures of IgG molecules. *MAbs* **4**, 17-23 (2012).
11. Barrios, Y., Jirholt, P. & Ohlin, M. Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor. *J. Mol. Recognit.* **17**, 332-338 (2004).
12. Kuroda, D., Shirai, H., Kobori, M. & Nakamura, H. Structural classification of CDR-H3 revisited: A lesson in antibody modeling. *Proteins: Structure, Function, and Bioinformatics* **73**, 608-620 (2008).
13. Fanning, L.J., Connor, A.M. & Wu, G.E. Development of the Immunoglobulin Repertoire. *Clinical Immunology and Immunopathology* **79**, 1-14 (1996).
14. Georgiou, G. et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32**, 158-168 (2014).
15. Lavinder, J.J., Horton, A.P., Georgiou, G. & Ippolito, G.C. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Curr. Opin. Chem. Biol.* **24**, 112-120 (2015).
16. Altelaar, A.F.M., Munoz, J. & Heck, A.J.R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* **14**, 35-48 (2013).
17. Bandeira, N., Pham, V., Pevzner, P., Arnott, D. & Lill, J.R. Automated de novo protein sequencing of monoclonal antibodies. *Nat. Biotechnol.* **26**, 1336-1338 (2008).
18. Guthals, A. et al. De Novo MS/MS Sequencing of Native Human Antibodies. *J. Proteome Res.* **16**, 45-54 (2017).
19. Tran, N.H. et al. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **6**, 31730 (2016).
20. Sen, K.I. et al. Automated Antibody De Novo Sequencing and Its Utility in Biopharmaceutical Discovery. *J. Am. Soc. Mass. Spectrom.* **28**, 803-810 (2017).
21. Fornelli, L. et al. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **90**, 8421-8429 (2018).
22. Zubarev, R.A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **15**, 12-16 (2004).
23. Taouatas, N., Drugan, M.M., Heck, A.J.R. & Mohammed, S. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* **5**, 405-407 (2008).
24. Lermyte, F., Valkenborg, D., Loo, J.A. & Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* **37**, 750-771 (2018).
25. Mao, Y., Valeja, S.G., Rouse, J.C., Hendrickson, C.L. & Marshall, A.G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **85**, 4239-4246 (2013).
26. Shaw, J.B. et al. Sequencing Grade Tandem Mass Spectrometry for Top–Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Anal. Chem.* **90**, 10819-10827 (2018).
27. Shaw, J.B. et al. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **92**, 766-773 (2020).
28. Teeling, J.L. et al. Characterization of new human CD20 monoclonal antibodies with potent cytolytic activity against non-Hodgkin lymphomas. *Blood* **104**, 1793-1800 (2004).
29. White, K.D., Frank, M.B., Foundling, S. & Waxman, F.J. Effect of immunoglobulin variable region structure on C3b and C4b deposition. *Mol. Immunol.* **33**, 759-768 (1996).
30. Ugurlar, D. et al. Structures of C1-IgG1 provide insights into how danger pattern recognition activates complement. *Science* **359**, 794-797 (2018).
31. Bondt, A. et al. Immunoglobulin G (IgG) Fab Glycosylation Analysis Using a New Mass Spectrometric High-throughput Profiling Method Reveals Pregnancy-associated Changes*. *Mol. Cell. Proteomics* **13**, 3029-3039 (2014).
32. Fort, K.L. et al. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **17**, 926-933 (2018).
33. Pacholarz, K.J. et al. Molecular Insights into the Thermal Stability of mAbs with Variable-Temperature Ion-Mobility Mass Spectrometry. *ChemBioChem* **17**, 46-51 (2016).
34. Chambers, M.C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918-920 (2012).

35. Klein, J., Carvalho, L. & Zaia, J. Application of network smoothing to glycan LC-MS profiling. *Bioinformatics* **34**, 3511-3518 (2018).
36. Lukauskas, J.K.h.S. (Zenodo, 2019).
37. Gibbons, C.W.B. (Computational Mass Spectrometry @ Pacific Northwest National Laboratory, 2020).
38. Park, J. et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **14**, 909-914 (2017).
39. Martí-Renom, M.A. et al. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure* **29**, 291-325 (2000).
40. Fiser, A. & Šali, A. in Methods in Enzymology, Vol. 374 461-491 (Academic Press, 2003).
41. Hansen, K. et al. A Mass-Spectrometry-Based Modelling Workflow for Accurate Prediction of IgG Antibody Conformations in the Gas Phase. *Angew. Chem. Int. Ed.* **57**, 17194-17199 (2018).
42. Saphire, E.O. et al. Crystal Structure of a Neutralizing Human IgG Against HIV-1: A Template for Vaccine Design. *Science* **293**, 1155-1159 (2001).
43. Harris, L.J., Larson, S.B., Hasel, K.W. & McPherson, A. Refined Structure of an Intact IgG2a Monoclonal Antibody. *Biochemistry* **36**, 1581-1597 (1997).
44. Arakawa, T. et al. Crystal structure of the anion exchanger domain of human erythrocyte band 3. *Science* **350**, 680-684 (2015).
45. Scapin, G. et al. Structure of full-length human anti-PD1 therapeutic IgG4 antibody pembrolizumab. *Nat. Struct. Mol. Biol.* **22**, 953-958 (2015).
46. Russel, D. et al. Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biol.* **10**, e1001244 (2012).
47. Marklund, Erik G., Degiacomi, Matteo T., Robinson, Carol V., Baldwin, Andrew J. & Benesch, Justin L.P. Collision Cross Sections for Structural Proteomics. *Structure* **23**, 791-799 (2015).
48. DeLano, W. (Schrödinger LLC, 2020).
49. Spoerry, C., Seele, J., Valentin-Weigand, P., Baums, C.G. & von Pawel-Rammingen, U. Identification and Characterization of IgdE, a Novel IgG-degrading Protease of Streptococcus suis with Unique Specificity for Porcine IgG*. *J. Biol. Chem.* **291**, 7915-7925 (2016).
50. von Pawel-Rammingen, U., Johansson, B.P. & Björck, L. IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *The EMBO Journal* **21**, 1607-1615 (2002).
51. Zubarev, R.A., Kelleher, N.L. & McLafferty, F.W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* **120**, 3265-3266 (1998).
52. Zubarev, R.A. et al. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **72**, 563-573 (2000).
53. Syrstad, E.A. & Turecček, F. Toward a general mechanism of electron capture dissociation. *J. Am. Soc. Mass. Spectrom.* **16**, 208-224 (2005).
54. Pacholarz, K.J. et al. Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling. *Angew. Chem. Int. Ed.* **53**, 7765-7769 (2014).
55. Hui, G.K. et al. The solution structure of the human IgG2 subclass is distinct from those for human IgG1 and IgG4 providing an explanation for their discrete functions. *J. Biol. Chem.* **294**, 10789-10806 (2019).
56. He, L. et al. Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS. *J. Am. Soc. Mass. Spectrom.* **28**, 827-838 (2017).

Supporting Information

Supplementary Figures and Tables

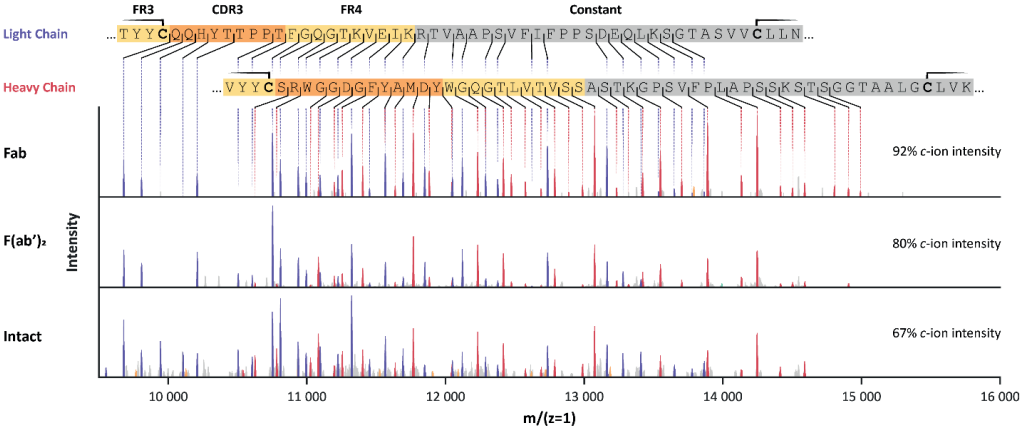


Figure S1 | Color-annotated charge-deconvoluted ECD-MS spectra of the intact, F(ab')$_2$ and Fab molecules of Trastuzumab showing primarily c-ion ladders covering the CDR3s of the LC (purple) and HC (red). Fragment *z*-ions originating from the LC and HC are indicated in green and orange respectively. The protein sequences are indicated above with the cysteine residues involved in the disulfide bonds denoted with a solid black line. Ion intensity within the shown mass window that can be explained by c-ions is highest for the Fab preparation, while considerably lower for the intact IgG1.

Table S1 | Some of the advantages and disadvantages of using native intact, F(ab')$_2$ or Fab molecules as precursor ions in the ECD-MS analysis.

Quality	Intact	F(ab')$_2$	Fab
Mass separation of *c*-ion fragments from precursor ions	Excellent	Excellent	Overlap with precursor ions
Spectral purity in terms of *c*-ions	Moderate due to interference from Fc *z*-ions	High	Highest
Ability for reproducible sample preparation	Excellent	Excellent	IgG1: Excellent IgG2-4: Poor due to use of nonspecific proteases
Sample homogeneity	Moderate due to glycosylation	Excellent	IgG1: Excellent IgG2-4: Poor due to use of nonspecific proteases

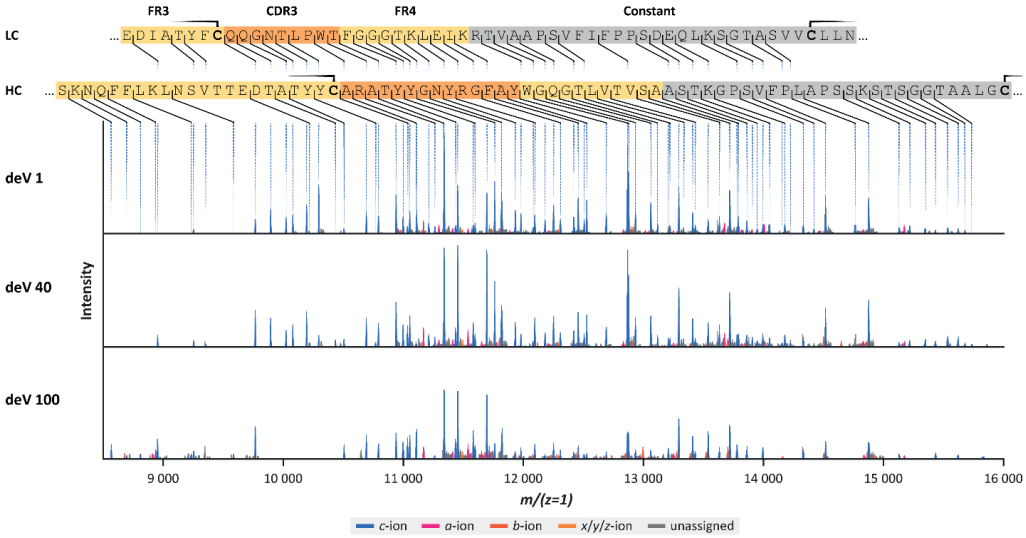


Figure S2 | Color annotated EChcD-MS spectra of the F(ab')$_2$ molecule of anti-DNP IgG1 measured by using different levels of supplemental collisional activation. Fragment *c*-ions of the LC and HC are indicated in blue, while *a*- and *b*-ions are indicated in purple and red respectively. C-terminal fragment ions (*z*/*y*/*z*-ions) are shown in orange. The sequences are indicated above with the cysteine residues involved in the disulfide bonds denoted with a solid black line. Although *c*-ion sequence coverage is somewhat extended by supplemental collisional activation, this is in less informative parts of the IgG, such as the FR3 and constant regions. Moreover, with the increase in supplemental collisional activation, more fragments of additional ion series are introduced, crowding the spectrum. For this experiment, the fragment ion search was performed with a 5-ppm tolerance.

Table S2 | Percentage of the total ion intensity within the 8500 < Mw < 16,000 Da window of Figure S2 that can be explained by searching for all different ion types with a 5-ppm tolerance. When minimizing collisional activation using a HCD direct eV setting of 1, about 70% of the total ion signal corresponds to *c*-ions. Increasing the level of collisional activation leads to a decrease in *c*-ion intensity, while *b*-ion intensity increases.

HCD deV	*a*-ions	*b*-ions	*c*-ions	*x* or *x*·-ions	*y*-ions	*z* or *z*·-ions
1	4.3	0.7	70.8	4.3	7.3*	1.6
40	3.2	2.3	69.4	3.0	6.0*	2.7
100	3.4	6.1	66.8	4.9	2.5*	1.7

* The relatively high *y*-ion intensity results from two LC *y*-ions that can also be explained as *c*-ions from the HC.

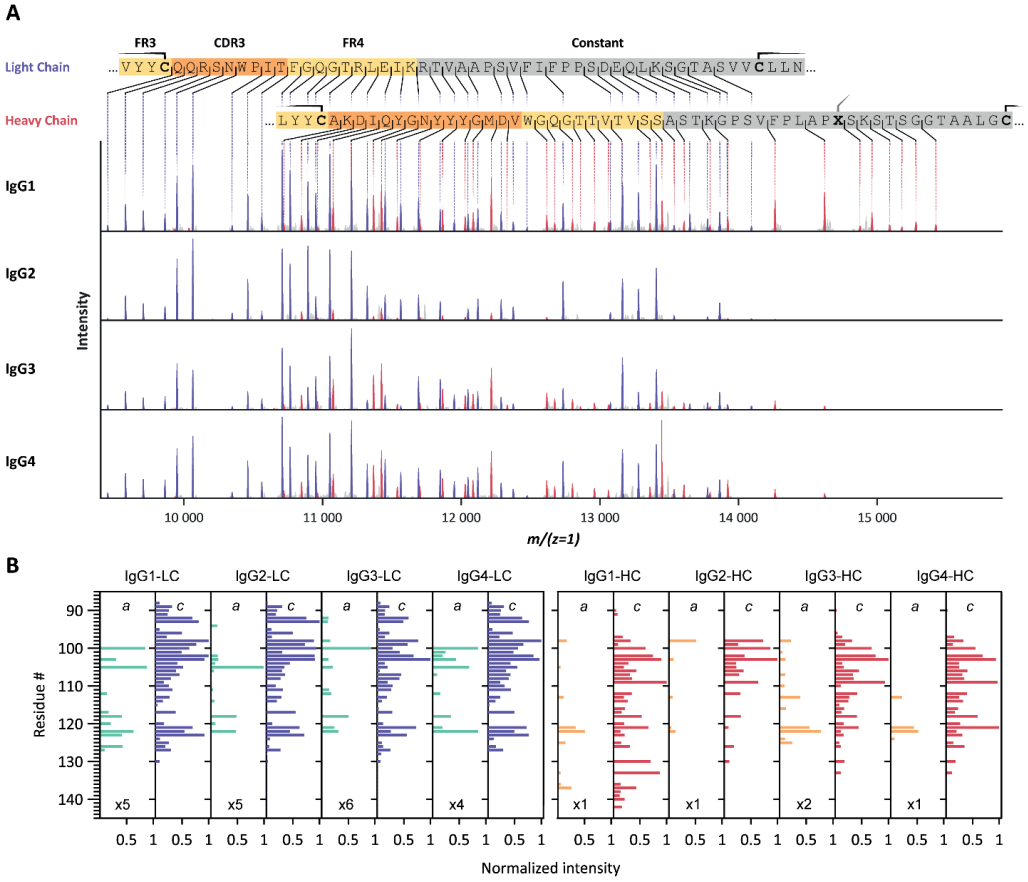


Figure S3 | Pure ECD of anti-CD20 IgG F(ab')$_2$ molecules produces clean CDR3-spanning *c*-ion ladders. **(A)** Charge-deconvoluted ECD-MS spectra of IgG1, IgG2, IgG3 and IgG4 F(ab')$_2$ molecules of anti-CD20 showing exclusively *c*-ion ladders covering the LC (purple) and HC (red) CDR3 regions. Cysteine residues involved in disulfide bonds are highlighted on the color-coded sequences by a solid line. Residue X, in …PLAPXS… in the heavy chain, represents a Ser for IgG1, and a Cys for IgG2-4. **(B)** Bar-chart displaying normalized *c*- and *a*-ion fragment intensities for each residue of IgG1, IgG2, IgG3 and IgG4 LC and HC. Although all ECD spectra for a given mAb are very similar, IgG1 HC *c*-ion fragments further extend towards the C-terminus. For some cleavage sites, a-ions complement missing *c*-ions.

Table S3 | Percentage of sequence-describing cleavages covered by *c*-ions in the ECD-MS spectra related to Figure 2 and Figure S3 resulting from a search with a 3-ppm tolerance. Near-complete coverage is obtained for the CDR3 and FR4 regions of both the LC and HC. The percentage of sequence-describing cleavages can be extended by using EChcD (ECD with supplemental collisional activation) as depicted in **Figure S2**. Opting for EChcD is, however, only recommended to complete sequence reads due to the formation of additional ion types (*e.g.*, *b/y*), hindering *de novo* analysis, as discussed in the main text.

mAb	LC CDR3	HC CDR3	LC CDR3-FR4	HC CDR3-FR4	LC	HC
IgG1 anti-CD20	**90.0**	**75.0**	95.0	77.8	21.0	15.4
IgG1 anti-DNP	**80.0**	**93.3**	85.0	92.3	16.4	16.7
IgG2 anti-CD20	**90.0**	**62.5**	95.0	48.1	18.7	6.3
IgG2 anti-DNP	**90.0**	**86.7**	95.0	80.8	17.3	11.0
IgG3 anti-CD20	**90.0**	**87.5**	95.0	88.9	21.5	10.8
IgG3 anti-DNP	**80.0**	**86.7**	85.0	84.6	14.5	11.1
IgG4 anti-CD20	**90.0**	**75.0**	95.0	77.8	20.6	11.3
IgG4 anti-DNP	**80.0**	**86.7**	85.0	84.6	15.9	12.2
Average	**86.3**	**81.7**	91.3	79.4	18.2	11.8

Table S4 | Percentage of the total ion intensity within the 9000 < Mw < 16,000 Da window of Figures 2A and S3 that can be explained by searching for all different ion types with a 3-ppm tolerance. Measurements were performed under minimal pressure and with an HCD direct eV setting of 1 to prevent supplemental collisional activation. Notably, in most ECD spectra typically quite a few of the ion signals cannot be explained, and therefore the fact that we can assign about above 80% of all fragment ion intensities is quite exceptional. The additional 20% is either chemical noise or other non-annotated ion signals (*e.g.*, small neutral losses of *c*-ions).

mAb	*a*-ions	*b*-ions	*c*-ions	*x* or *x*·-ions	*y*-ions	*z* or *z*·-ions
IgG1 anti-CD20	**5.5**	1.4	**76.0**	-	0.2	0.3
IgG1 anti-DNP	**3.9**	1.0	**73.6**	0.3	6.2*	0.5
IgG2 anti-CD20	**2.2**	0.3	**83.9**	0.1	-	-
IgG2 anti-DNP	**2.8**	0.3	**73.9**	2.4	2.0	0.7
IgG3 anti-CD20	**4.2**	1.2	**78.1**	-	-	-
IgG3 anti-DNP	**1.9**	1.4	**73.2**	1.2	0.9	-
IgG4 anti-CD20	**5.3**	0.7	**80.0**	-	-	-
IgG4 anti-DNP	**1.8**	0.5	**70.3**	1.5	1.5	0.7
Average	**3.4**	0.9	**76.1**	0.7	1.3	0.2

* The relatively high *y*-ion intensity results from two LC *y*-ions that can also be explained as *c*-ions from the HC.

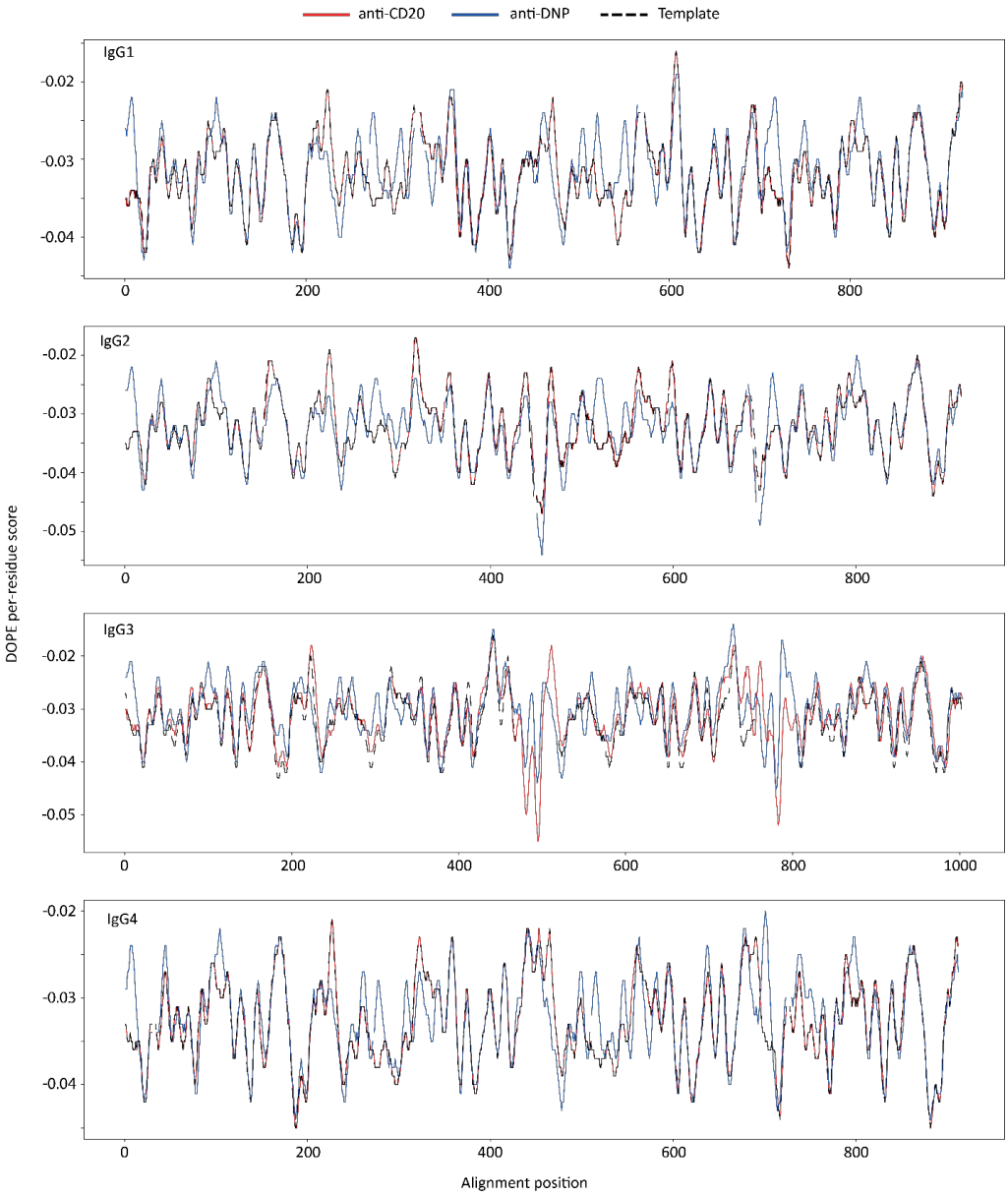


Figure S4 | Discrete Optimized Protein Energy (DOPE) profiles showing the DOPE per-residue score as function of the residue number (alignment position) for F(ab')$_2$ subclasses (IgG 1-4) comparatively modelled for anti-CD20 and anti-DNP. While the total DOPE score displays an invert correlation with the native character of a structure – the total DOPE score is lower for more native like (in vacuo) structures[1] –, the DOPE score of each residue, displayed here, indicates how much a residue contributes to the protein energy as well as, upon comparison with reference values, how much the local structure potentially differs from a template or analogue "ideal" structure. Comparison of the three curves sets (black dash, red, and blue) highlights the perfect match between anti-CD20 (red) and the templates (black dashes) while anti-DNP (blue) deviates slightly more. The rather small positive deviations of the blue curve compared to the template nevertheless indicate that we are close to optimal structures (following the optimization of flexible loops of residues). The templates used are IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3.

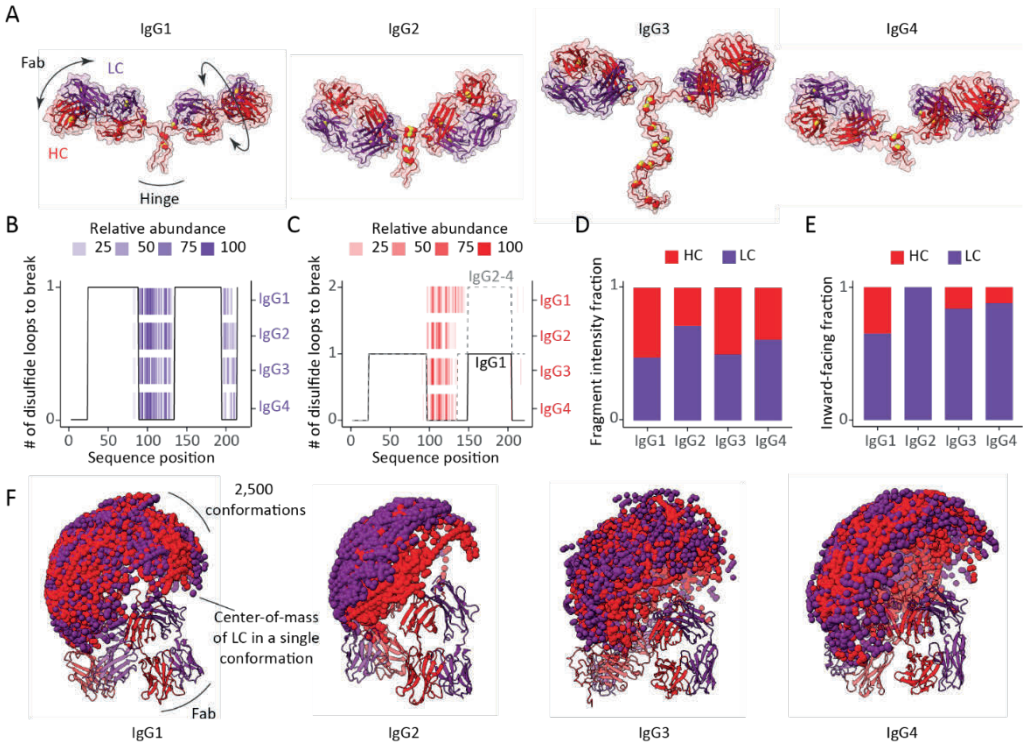


Figure S5 | Influence of structural constraints on ECD fragmentation of F(ab')$_2$ from different subclasses of anti-DNP IgG. (A) Structures of F(ab')$_2$ comparatively modelled against available structures of IgG subclasses (IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3). Disulfide bridges are displayed as paired yellow spheres. **(B, C)** Number of disulfide bonds that span each sequence position of (**B**) the LC and (**C**) the HC in different subclasses of IgG anti-DNP overlaid with the corresponding ECD fragments. **(D)** Fractions of ECD fragment intensities originating from the LC and HC for the different subclasses of IgG anti-DNP. **(E)** Fractions of exposed LC and HC in the sampling of 500 most compact conformations. See **Figure S7** for a more detailed description of the estimated chain relative exposure based on the center-of-mass calculations. **(F)** Sampling of F(ab') arm movements displayed by spheres representing the center-of-mass of (purple) LCs and (red) HCs, showing remarkable differences in between the different subclasses.

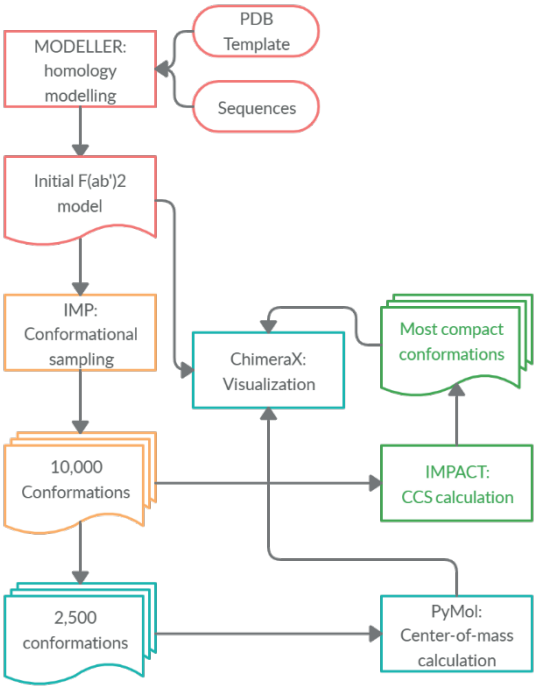


Figure S6 | Modelling workflow used to obtain F(ab')$_2$ structural models for of anti-CD20 and anti-DNP IgG1, IgG2, IgG3 and IgG4. First, the structures were homology modelled against the available structures of IgGs from the different subclasses (taken were the following PDB structures; IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3). Next, the F(ab') arm movement was sampled by using the Integrative Modelling Platform rapidly exploring the random tree algorithm[2], where the F(ab') portions were set as a rigid body and amino acids in the hinge region were set to rotate in a random manner in line with their torsional space. From the 10,000 generated conformations 2500 were used for the calculation of the center-of-mass (COM) with PyMol for both the entire F(ab')$_2$ and each chain separately. For COM calculations of the IgG3 F(ab')$_2$, we included only the first two hinge disulfides to avoid the offset created by the large hinge region. Subsequently the average distances from the heavy or light chain's COM to the full molecule's COM were used to estimate the relative exposure of the light and heavy chain within each conformation. In parallel, the entire population of F(ab')$_2$ conformations was subject to calculation of collisional cross-sections (CCS) making use of the Impact software[3]. All visualizations were performed in ChimeraX environment[4].

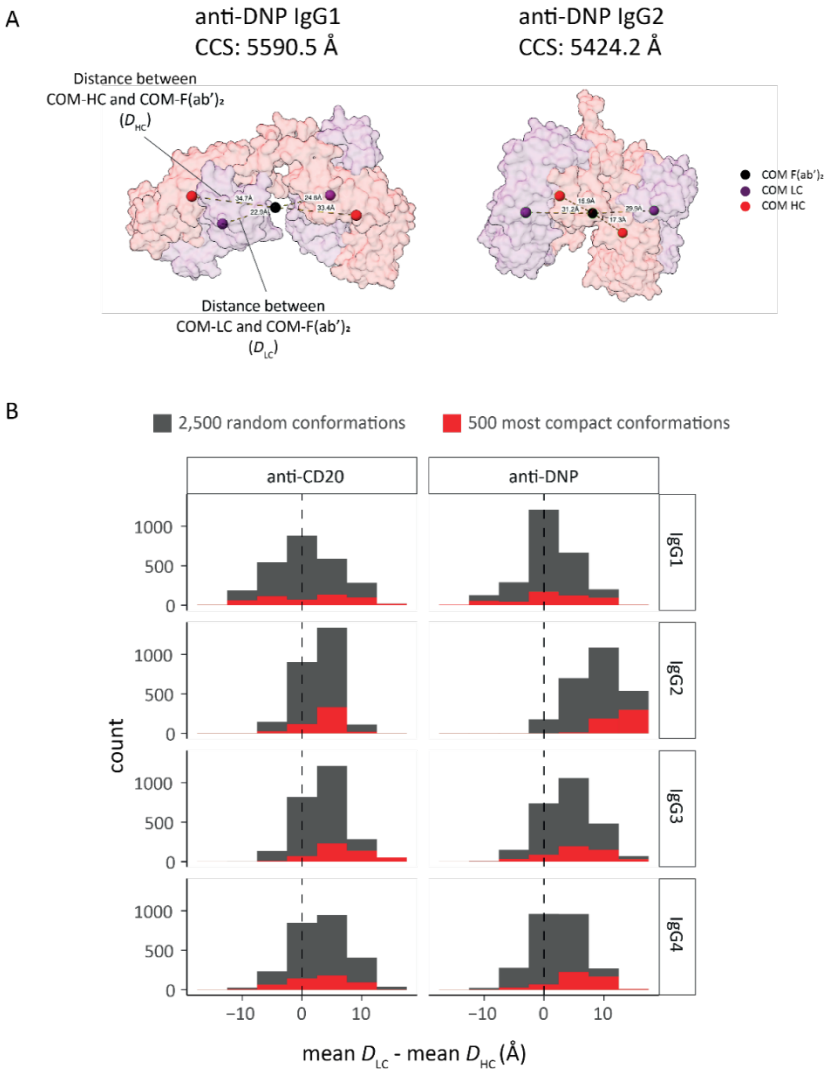


Figure S7 | Estimation of the relative exposure in the F(ab')$_2$ molecules based on the Center-of-mass (COM) based calculations (Figure S6). (A) Distances between the light chain COM and F(ab')$_2$ COM (D_{LC}) and the heavy chain COM and F(ab')$_2$ COM (D_{HC}) are visualized in compact conformations of anti-DNP IgG1 and IgG2. **(B)** Distributions of differences between average Center-of-mass D_{LC} and D_{HC} in 2500 random and 500 compact conformations of anti-CD20 and anti-DNP IgG1, IgG2. IgG3 and IgG4.

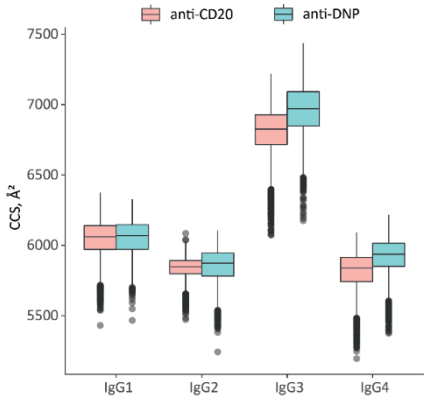


Figure S8 | Distribution of collisional cross-sections of IgG anti-CD20 and IgG anti-DNP conformations generated by using the Integrative Modelling Platform[2]. Collisional cross-sections were calculated with Impact software[3].

Table S5 | Amino acid sequences of the anti-CD20 and anti-DNP IgG mAbs used in our experiments. The cleavage site of the IdeS enzyme that produces the F(ab')$_2$ is indicated with a bold forward slash in the heavy chain sequences.

IgG anti-CD20 (7D8)

IgG1 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRFTISRDNAKNSLYLQMNSLR
AEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHT
FPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLG/GPSVFLFPPKPKDTLMISR
TPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQ
PREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMH
EALHNHYTQKSLSLSPGK

IgG2 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRFTISRDNAKNSLYLQMNSLR
AEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHT
FPAVLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVERKCCVECPPCPAPPVA/GPSVFLFPPKPKDTLMISRTPEV
TCVVVDVSHEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREP
QVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALH
NHYTQKSLSLSPGK

IgG3 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRFTISRDNAKNSLYLQMNSLR
AEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHT
FPAVLQSSGLYSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELKTPLGDTTHTCPRCPEPKSCDTPPPCPRCPEPKSCDTPP
PCPRCPEPKSCDTPPPCPRCPAPELLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNAKTKPREEQ
YNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWE
SSGQPENNYNTTPPMLDSDGSFFLYSKLTVDKSRWQQGNIFSCSVMHEALHNRFTQKSLSLSPGK

IgG4 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRFTISRDNAKNSLYLQMNSLR
AEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHT
FPAVLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYGPPCPSCPAPEFLG/GPSVFLFPPKPKDTLMISRTPE
VTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPRE
PQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEAL
HNHYTQKSLSLSLGK

κ Light Chain

EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLTISSLEPEDFAVYY
CQQRSNWPITFGQGTRLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYS
LSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

IgG anti-DNP (G2a2)

IgG1 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRISITRDTSKNQFFLKLNSVT
TEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTF
PAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLG/GPSVFLFPPKPKDTLMISRT
PEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP
REPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHE
ALHNHYTQKSLSLSPGK

IgG2 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRISITRDTSKNQFFLKLNSVT
TEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTF
PAVLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVERKCCVECPPCPAPPVA/GPSVFLFPPKPKDTLMISRTPEVT
CVVVDVSHEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREPQ
VYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHN
HYTQKSLSLSPGK

IgG anti-DNP (G2a2) (continued)

IgG3 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRISITRDTSKNQFFLKLNSVT
TEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTF
PAVLQSSGLYSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELKTPLGDTTHTCPRCPEPKSCDTPPPCPRCPEPKSCDTPPP
CPRCPEPKSCDTPPPCPRCPAPELLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNAKTKPREEQY
NSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWES
SGQPENNYNTTPPMLDSDGSFFLYSKLTVDKSRWQQGNIFSCSVMHEALHNRFTQKSLSLSPGK

IgG4 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRISITRDTSKNQFFLKLNSVT
TEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTF
PAVLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYGPPCPSCPAPEFLG/GPSVFLFPPKPKDTLMISRTPEV
TCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREP
QVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALH
NHYTQKSLSLSLGK

κ Light Chain

DIRMTQTTSSLSASLGDRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYF
CQQGNTLPWTFGGGTKLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYS
LSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Table S6 | Overview of the molecular masses of the used intact IgGs, F(ab')$_2$ and Fabs. The theoretical mass is obtained from the provided protein sequences as shown in **Table S3**, the experimental mass is determined by mass spectrometry.

Species	$MW_{theoretical}$ (Da)	$MW_{experimental}$ (Da)	Mass difference (Da)
Trastuzumab G1F/G0F	148,476	148,220	-256.0*
Trastuzumab F(ab')$_2$	97,629	97,631	1.6
Trastuzumab Fab	47,500	47,502	1.9
IgG1 anti-DNP F(ab')$_2$	99,201	99,203	2.3
IgG2 anti-DNP F(ab')$_2$	98,630	98,632	1.7
IgG3 anti-DNP F(ab')$_2$	109,352	109,355	3.2
IgG4 anti-DNP F(ab')$_2$	98,782	98,783	1.2
IgG1 anti-CD20 F(ab')$_2$	98,430	98,432	2.3
IgG2 anti-CD20 F(ab')$_2$	97,859	97,862	2.3
IgG3 anti-CD20 F(ab')$_2$	108,582	108,588	6.0
IgG4 anti-CD20 F(ab')$_2$	98,011	98,014	2.7

* Corresponding to the mass shift induced by C-terminal lysine clipping of both HCs.

Supplementary References

1. Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507-2524 (2006).
2. Russel, D. et al. Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biol.* **10**, e1001244 (2012).
3. Marklund, Erik G., Degiacomi, Matteo T., Robinson, Carol V., Baldwin, Andrew J. & Benesch, Justin L.P. Collision Cross Sections for Structural Proteomics. *Structure* **23**, 791-799 (2015).
4. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14-25 (2018).

CHAPTER 8

GENERATING INFORMATIVE SEQUENCE TAGS FROM ANTIGEN-BINDING REGIONS OF HEAVILY GLYCOSYLATED IGA1 ANTIBODIES BY NATIVE TOP-DOWN ELECTRON CAPTURE DISSOCIATION

Jean-Francois Greisch[1,2,*], Maurits A. den Boer[1,2,*], Frank Beurskens[3], Janine Schuurman[3], Sem Tamara[1,2], Albert Bondt[1,2], and Albert J.R. Heck[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Genmab, Uppsalalaan 15, 3584 CT Utrecht, The Netherlands

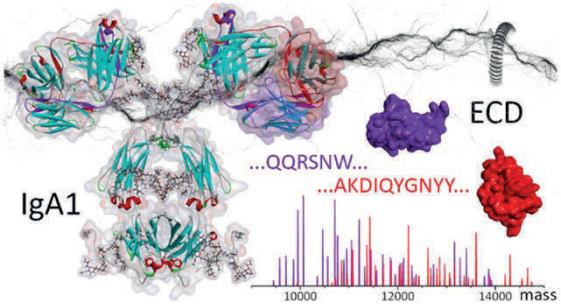
[*] These authors contributed equally

Published as:

Generating Informative Sequence Tags from Antigen-Binding Regions of Heavily Glycosylated IgA1 Antibodies by Native Top-Down Electron Capture Dissociation
Jean-Francois Greisch, Maurits A. den Boer, Frank Beurskens, Janine Schuurman, Sem Tamara, Albert Bondt, and Albert J. R. Heck
Journal of the American Society for Mass Spectrometry 2021, 32, 6, 1326-1335
DOI: 10.1021/jasms.0c00461

Abstract

Immunoglobulins A (IgA) include some of the most abundant human antibodies and play an important role in defending mucosal surfaces against pathogens. The unique structural features of the heavy chain of IgA subclasses (called IgA1 and IgA2) enable them to polymerize via the joining J-chain, resulting in IgA dimers but also higher oligomers. While secretory sIgA oligomers are dominant in milk and saliva, IgAs exist primarily as monomers in serum. No method currently allows disentangling the millions of unique IgAs potentially present in the human antibody repertoire. Obtaining unambiguous sequence reads of their hypervariable antigen-binding regions is a prerequisite for IgA identification. We here report a mass spectrometric method that uses electron capture dissociation (ECD) to produce straightforward-to-read sequence ladders of the variable parts of both the light and heavy chains of IgA1s, in particular, of the functionally critical CDR3 regions. We directly compare the native top-down ECD spectra of a heavily and heterogeneously *N*- and *O*-glycosylated anti-CD20 IgA1, the corresponding *N*-glycosylated anti-CD20 IgG1, and their Fab parts. We show that while featuring very different MS1 spectra, the native top-down ECD MS2 spectra of all four species are nearly identical, with cleavages occurring specifically within the CDR3 and FR4 regions of both the heavy and light chains. From the sequence-informative ECD data of an intact glycosylated IgA1, we foresee that native top-down ECD will become a valuable complementary tool for the de novo sequencing of IgA1s from milk, saliva, or serum.



Introduction

Over the past decades, IgG-based antibodies have been increasingly used in the clinic in the fields of oncology, hematology, autoimmune diseases, and infections[1]. It is, however, becoming increasingly apparent that other immunoglobulins, *e.g.*, immunoglobulin A (IgA), might provide a useful alternative[2].

In humans, the production of IgAs surpasses that of all the other Ig classes combined. While most of the daily IgA production (estimated at 60 mg per kilogram of body weight) is located at mucosal surfaces, offering frontline protection against invading pathogens, 2-3 mg/mL is also present in serum[3]. The action of IgAs is characterized by their ability to recruit different effector cells, *i.e.*, polymorphonuclear cells or neutrophils, as well as to activate monocytes and macrophages. Not surprisingly, there is, therefore, a growing interest in IgA-based therapeutics[3].

In terms of sequence and structure, recombinant monomeric IgG and IgA molecules feature quite a few similarities. They both consist of two heavy and two light chains connected by disulfide bridges, and both harbor a similar highly constant part (Fc) and a variable antigen-binding part (Fab) that contains the complementarity-determining regions (CDRs). The light chains (LC) can even be identical. IgGs and IgAs nevertheless also exhibit some very distinctive subclass-dependent features. The heavy chain (HC) constant domain of human IgGs is divided into four subclasses, namely, IgG1, IgG2, IgG3, and IgG4, which share between 83 and 96% of their sequence. Additionally, IgG subclasses exhibit distinctive disulfide bridging patterns characterized by different numbers of disulfide bonds in the hinge region and different positioning of disulfides between LC and HC[4]. For human IgA, two subclasses are recognized, namely, IgA1 and IgA2. Compared to IgGs, both IgA1 and IgA2 carry more cysteines on their HC, resulting in a higher number of interchain disulfide bridges[3]. The IgA subclasses primarily differ in their hinge region, with IgA1 containing an additional stretch of 13 residues between the Fab and Fc regions. Distinctively, the IgA1 HC hinges are rich in serine/threonine residues (as well as proline residues) and can bear up to six core 1 *O*-glycans. For the larger number of *N*-glycans carried by IgA1s compared to IgGs, it further adds to their structural complexity[4]. Consequently, the higher degree of molecular heterogeneity resulting from the coexistence of multiple glyco-proteoforms is more challenging to tackle analytically for IgAs than for IgGs. This difference in structural heterogeneity becomes quite apparent when analyzing intact IgG1s and IgA1s by mass spectrometry (MS) (**Figure 1**).

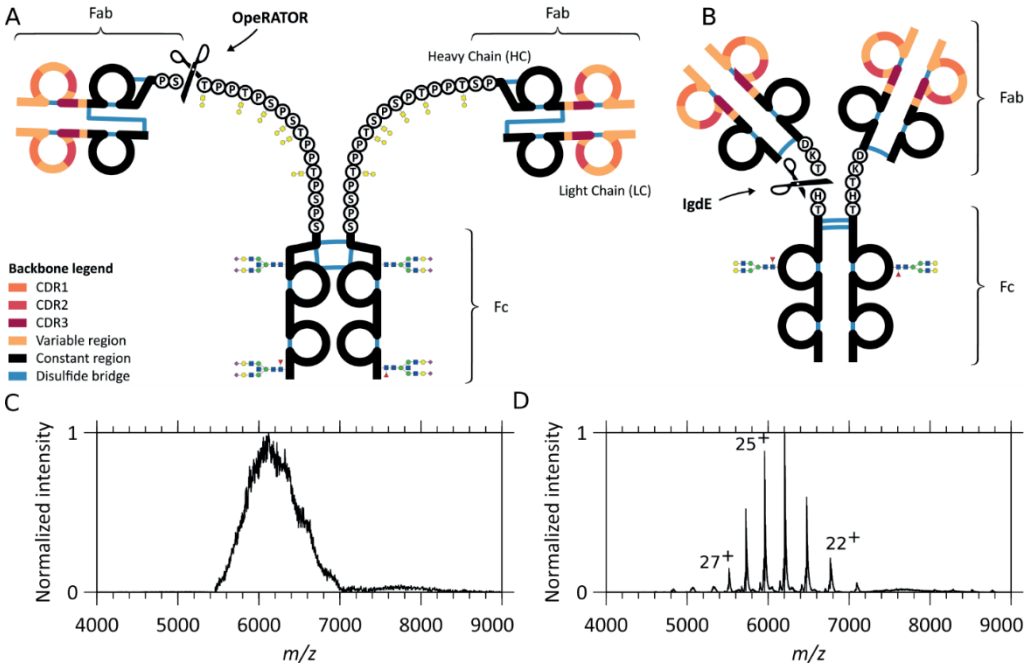


Figure 1 | Schematic overview of similarities and differences in IgA1 and IgG1 and the resulting native mass spectra. Comparison of **(A)** IgA1 and **(B)** IgG1 structures: the two variable regions (colored) display three antigen-binding CDRs each and one constant region (black). Notably, although the light and heavy chain are connected by a single disulfide bond in both IgA1 and IgG1, the connectivity is different. CDR-containing Fab fragments of either IgA1s or IgG1s can be separated from the glycosylated constant Fc portion by proteolytic cleavage (scissors). The OpeRATOR enzyme can cleave IgA1 N-terminally of all *O*-glycosylation sites, producing predominantly Fab molecules terminated by Ser105 on the constant region of the HC (scissors), whereas IgdE cleaves IgG1 at one specific sequence motif. While IgA1 HCs bind to LCs via a more N-terminal HC cysteine than for IgG1, they also differ by the number and location of intrachain disulfides: 6 for IgA1s against 4 for IgG1s. The IgA1 hinge region is extended compared to IgG1. Although the *O*-glycans are here depicted uncapped for simplicity, they can be variably extended by additional Gal and NeuNAc residues. Native MS1 spectra of **(C)** intact anti-CD20 IgA1 and **(D)** anti-CD20 IgG1. While baseline-resolved ion signals can be detected for IgG1, the structural heterogeneity of IgA1 leads to charge-unresolved ion signals, hampering direct mass determination.

Accurate sequence determination, a major prerequisite for the ultimate goal of clinical application of antibodies, requires high-quality, streamlined data sets. Mass spectrometry is thereby often the method of choice to obtain sequence information. In mass-spectrometry-based sequencing, forming ions from a single fragment pair (a/x, b/y, or c/z) can be beneficial for the de novo sequencing of antibodies as it avoids spectral congestion and dilution of signal intensity in fragmentation spectra. While conventional mass spectrometric methods such as CID or IRMPD can, in principle, be optimized to enhance the abundance of a single fragment pair, they often come with limited sequence coverage, extensive side-chain cleavages, and substantial internal fragmentation. These disadvantages make electron capture dissociation (ECD) an alternative choice[5-8]. Applied to denatured and native intact antibodies[9,10], ECD results in significant backbone cleavages in both the LC and HC variable regions, primarily leading to the formation of ($c/z\cdot$)

fragment ions. Furthermore, since intramolecular disulfide bridges of native proteins are not frequently cleaved during ECD[11], enhanced fragment formation is observed for sequence segments not involved in disulfide-bridged loops, *e.g.*, for the segments covering the LC and HC CDR3s[9,12].

Notwithstanding the tremendous structural complexity of the precursor molecule, we here present a method to obtain straightforward-to-read amino acid sequence ladders from the CDR3 of intact IgA1 molecules. We further optimize ECD toward the *de novo* sequencing of IgA1 CDR3s and their downstream regions, generating clean fragment ion series composed solely of *c*-ions[8]. Using native mass spectrometry conditions, we achieve maximal separation between the precursor in the high m/z range and the informative fragment ions in the lower m/z range. We also show that, although IgA1 Fab isolation reduces the complexity of the spectrum and simplifies the precursor ion selection, simplified precursors are not required to obtain straightforward sequence reads from IgA1 CDR3 regions. The method thus proves to be generally applicable to IgA1 immunoglobulins and their proteoforms. We conclude by addressing fragment intensity prediction in terms of the energy required to separate fragments stabilized by noncovalent interactions.

Experimental Section

Antibody Samples

Monoclonal IgAs and IgGs against CD20 (anti-CD20)[13] were recombinantly expressed and provided by Genmab (Utrecht, The Netherlands). Briefly, mAbs were expressed in Expi293F cells (Life Technologies, Waltham, MA, USA) by transient transfection using an ExpiFectamine 293 transfection kit (Life Technologies, Waltham, MA, USA) according to the instructions of the manufacturer. The culture was harvested 5 days post-transfection by centrifugation for 10 min at 3000 *g*, followed by filter sterilization of the supernatant using a 0.22 μm filter and storage at 4 °C. The amino acid sequences of these antibodies are provided in **Table S1**.

Preparation of Intact IgA and Fab Samples for Native Top-Down MS Analysis

Fab molecules were generated by overnight digestion at 37 °C of the IgAs bound to CaptureSelect IgA affinity matrix (Thermo Fisher Scientific) using 1 U/μg of the SialEXO sialidase cocktail and the OpeRATOR *O*-glycopeptidase from *Akkermansia muciniphila* (Genovis AB, Llund, Sweden) for IgA1s and on-bead digestion of the IgGs bound to CaptureSelect FcXL affinity matrix using 1 U/μg FabALACTICA IgdE (Genovis AB, Llund, Sweden) as described previously[14].

Prior to native top-down analysis, buffers were exchanged to aqueous 150 mM ammonium acetate (pH 7.5) through six consecutive dilution and concentration steps at 4 °C using Amicon Ultra centrifugal filters with a 10 kDa molecular weight cutoff (Merck KGaA, Darmstadt, Germany). IgA and IgG concentrations used for native electrospray ionization mass spectrometry were around 4 μM.

Native Top-Down ECD MS

Top-down MS of native immunoglobulins was performed on an ultrahigh mass range (UHMR) Q-Exactive Orbitrap (Thermo Fisher Scientific, Bremen, Germany) using an ECD cell developed by e-MSion[15]. First, intact or Fab precursor ions were mass-selected by the quadrupole. The in-source trapping desolvation voltage displays an optimum in terms of ECD at -100 V for Fabs and intact (glycosylated) IgAs and IgGs (data not shown), likely a consequence of the small structural changes occurring as the effective temperature of the ion is raised and leading to cleavage enhancement[16]. The ions were then transmitted into the ECD cell where they were subjected to electron capture dissociation. Second, upon transfer from the ECD cell to the HCD (high-energy C-trap dissociation) cell, post-ECD collisional activation was kept to a minimum (HCD direct eV setting = 1) to avoid (b/y) ion formation and concentrate the fragment signal intensity into the *c*- and *z*-ions of interest. Overall, HCD cell trapping and extraction parameters were optimized for low nitrogen collision gas pressures to achieve efficient detection of the subunits and fragments. All spectra were acquired with the noise threshold parameter set to 3.64. Whenever possible, a single charge state of the precursor (Fab or intact Ig) was isolated to simplify interpretation of the fragment ion spectra. All spectra were acquired at a set resolution of 200,000 at m/z 400.

Data Analysis

Processing of the fragmentation spectra involved the conversion of raw files to mzML format by Proteowizard[17]. We used the MSDeisotope python library (Joshua Klein, Boston University CBMS)[18,19] with a minimum_score = 10.0 and mass_error_tolerance = 0.02 to generate a charge-deconvoluted spectrum with all of the isotopic peaks retained[18,19].

Comparison of unprocessed data with the spectrum reconstructed from the charge-deconvoluted $m/[z = 1]$ spectra corresponding to fragments sharing the same charge state involved shifting each charge-deconvoluted spectrum by $(z - 1) \times 1.007276$, dividing them by their respective charge, *z*, and superposing the result.

Fragments were assigned by applying LcMsSpectator (Pacific Northwest National Laboratory)[20,21] to the charge-deconvoluted spectra generated by MSDeisotope. The accuracy threshold was set to ± 3 ppm for all assignments following recalibration of the fragment's m/z via shifting by the average error on assignable *c*-ion fragments. The results were exported as .tsv files for further analysis. Sequence assignment accommodated the major ECD ion types (c, z, $z\cdot$) without considering H_2O and NH_3 neutral losses, except when explicitly mentioned.

Spectral comparison of native top-down ECD spectra involved an in-house implementation of the cross-correlation approach pioneered for mass spectra by Yates and co-workers[22]. Cross-correlation provides a metric for the pairwise comparison of spectra. The approach is ideally suited for cases where fragment isotope envelopes expand well beyond the first isotopic peak, and this peak's intensity is diminished relative to the base peak in the envelope. For charge-deconvoluted spectra, we report the normalized cross-correlation values for a 1000 m/z mass window slid over the spectra in steps of 10 m/z, except when specified otherwise.

Interaction energies within the IgA1 structure were computed as follows. We used Modeller[23] to generate a homology model for the anti-CD20 IgA1 sequence using the PDB ID 3M8O IgA1 structure. The LC and HC backbone bonds between the cysteine residues involved in intrachain disulfide loops were systematically cleaved, leading to the generation of a separate model structure for each cleavage. The resulting structures were geometrically relaxed using foldX (v4)[24,25], and the interpolypeptide interactions were computed using the same software.

Results and Discussion

Comparison of IgG1 and IgA1 Glycosylation and Resulting Native Mass Spectra

IgG1s typically harbor a single *N*-glycosylation site in the Fc region of each of the two heavy chains. Their IgA1 counterparts display a much higher glycosylation and glycan complexity. *N*-Linked oligosaccharides contribute up to 6-7% of the mass of human IgA1[3,26], with *N*-glycans found attached to residues Asn263 and Asn459[3,27]. In terms of composition, the *N*-linked glycans of serum and secretory IgAs primarily are of the diantennary complex type, with a small fraction showing triantennary or more branched structures. Additional heterogeneity arises from variable fucosylation levels and numbers of galactoses and sialic acids at the branched termini[28-30]. IgA1s also harbor three to six core 1 and/or Tn *O*-linked glycans, composed principally of *N*-acetyl galactosamine, galactose, and zero, one, or two sialic acids attached to Ser and Thr sites in the IgA1 hinge region[29,30]. Furthermore, compared to IgG1, IgA1 molecules have an extended hinge region and feature a different disulfide bridging pattern, as schematically depicted in **Figure 1A** and **B**. To illustrate IgA1 heterogeneity, a comparison of the native mass spectra of intact IgA1 and IgG1 is displayed in **Figure 1C** and **D**. Whereas, for IgG1, baseline-resolved mass spectra allow even the direct analysis of the glycan composition[31], accurate mass determination is severely hampered for IgA1 due to the unresolvable charge states.

Impact of Glycosylation on the Native Top-Down ECD Mass Spectra of Intact IgA1 and IgG1

In **Figure 2**, we compare the native top-down ECD mass spectra of intact anti-CD20 IgA1 and IgG1 in order to assess the impact of the differences in structure and glycosylation on electron capture dissociation. These monoclonal antibodies, which possess identical LCs and identical HC variable regions, only differ in their HC constant regions. The simplicity of the anti-CD20 IgG1 MS1 spectrum allowed isolation of the single 23+ charge state (isolation window of 5 m/z). On the other hand, for the anti-CD20 IgA1, the absence of resolvable charge states (**Figure 1C**) led us to perform ECD on an m/z window encompassing the whole charge state distribution ($5700 < m/z < 6900$) as well as on an m/z window ($6190 < m/z < 6400$) englobing most of the IgA1 25+ charge state (isolation window of 210 m/z). Notably, for IgA1, the resulting ECD spectra looked nearly identical and were thus not influenced by isolating either the broader or narrower m/z window (**Figure S2**).

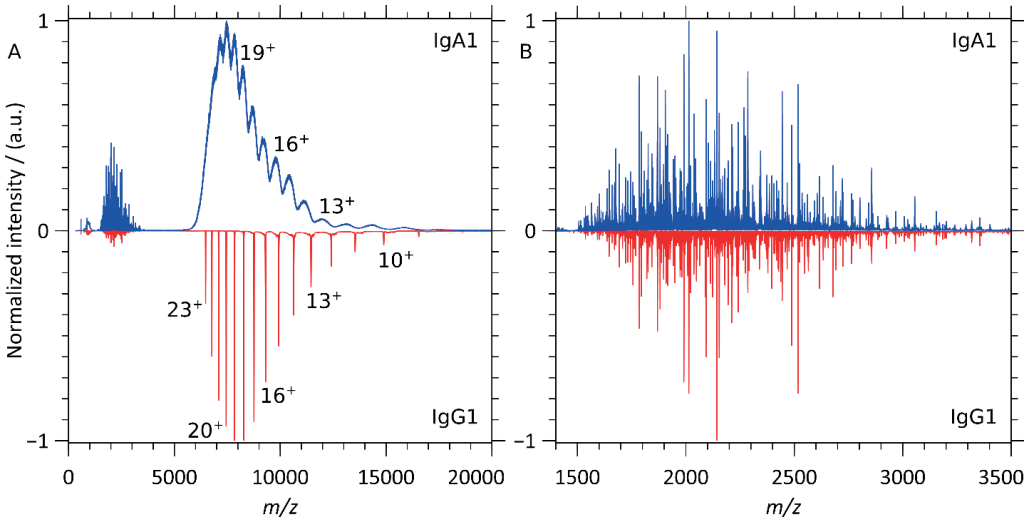


Figure 2 | Native top-down ECD MS of anti-CD20 IgA1 and IgG1. (A) ECD spectra of intact anti-CD20 IgA1 (blue) and IgG1 (red) dominated by signals resulting from successive electron capture without dissociation. **(B)** Lower *m/z* spectral regions corresponding to the single-chain ECD fragments of intact anti-CD20 IgA1 (blue) and IgG1 (red).

In the higher m/z range ($5000 < m/z < 20{,}000$), electron capture primarily yields charge-reduced peaks –ions absorbing from 1 to 15 electrons without incurring dissociation, *i.e.*, ECnoD[7] – resulting for IgG1 in a charge-resolved distribution of precursors. Similar ECnoD processes are observed for IgA1, and although the resulting charge states are still very broad, they improve upon the initial spectrum (**Figure 1C**). The detectable charge states can yield an average mass estimate for the IgA1 molecules, 157,660 ± 170 Da (**Table S1**).

Despite the large difference in the complexity of the precursor anti-CD20 IgA1 and IgG1 molecules, the low m/z ($1400 < m/z < 3500$), ECD fragments of these intact immunoglobulins are highly similar, as displayed in **Figure 2B** and quantified using cross-correlation analysis in the next section. The differences in structure and glycosylation, therefore, do not seem to impact ECD fragment formation for the CDR3s and FR4s of immunoglobulins.

Comparison of Intact and Fab Arm IgA1 and IgG1 ECD-Derived *c*-Ion Sequence Ladders

In order to compare fragment masses and abundances, it is more convenient to proceed by inspecting the charge-deconvoluted spectra. In **Figures 3** and **S1**, we first assessed the quality of our charge deconvolution process by comparing, for the intact anti-CD20 IgA1, the mass spectrum reconstructed from the charge-deconvoluted mass spectra of fragments segregated according to their charge state with the corresponding unprocessed ECD spectrum. Some large ions with low abundances and low charge states ($m/z > 3000$, $z = 4$) are absent from the charge-deconvoluted spectrum. This is a consequence of the

S/N and intensity criteria used for peaks to be recognized as part of a given isotopic distribution. Still, although not perfect, a cross-correlation analysis confirms the negligible loss of spectral information introduced by the charge deconvolution process.

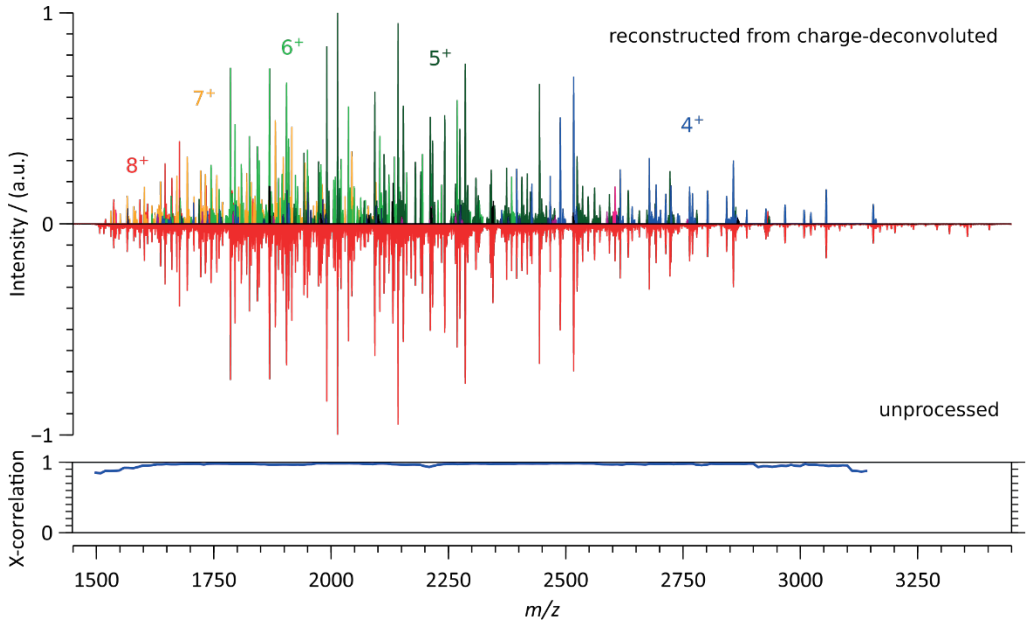


Figure 3 | Comparison of the low *m/z* region of the native top-down ECD spectra of intact anti-CD20 IgA1 spectrum reconstructed from charge-deconvoluted spectra corresponding to fragment charge states ranging from 1+ to 13+ (top), with the unprocessed fragment ion spectrum (bottom). The most represented charge states are color annotated. Cross-correlation analysis, using a 100 *m/z* sliding window in steps of 10 *m/z*, yields an average high score of 0.964.

Having established that the charge-deconvoluted spectra account for the majority of the fragment ion signal in the unprocessed spectra, we first compare the charge-deconvoluted ECD spectra of the intact anti-CD20 IgA1 and the anti-CD20 IgG1 (**Figure 4**). In both cases, straightforward-to-read *c*-ion ladders dominate the ECD spectrum in this region and enable sequence determination for both the LC and HC CDR3 and FR4. Close-to-identical sequence ladders are obtained for anti-CD20 IgA1 and IgG1 up to $m/[z = 1]$ 14,000 as established by cross-correlation analysis. As expected, from $m/[z = 1]$ 14,000, the IgA1 spectrum is shifted relatively to the IgG1 spectrum due to the mass difference between a Pro and a Gly residue in the Fd region (IgA.xxx**P**TSPK - IgG.xxxTK**G**PS ≡ 97.0528 Da - 57.0215 Da = 40.0313 Da). Additional differences in the higher $m/[z = 1]$ range can be attributed to differences in the disulfide bridging patterns as discussed further below. Practically, the substantial structural heterogeneity of the intact anti-CD20 IgA1 does not affect the ECD process, as determined from similar fragment ion spectra for IgA1 and IgG1 in the lower m/z region.

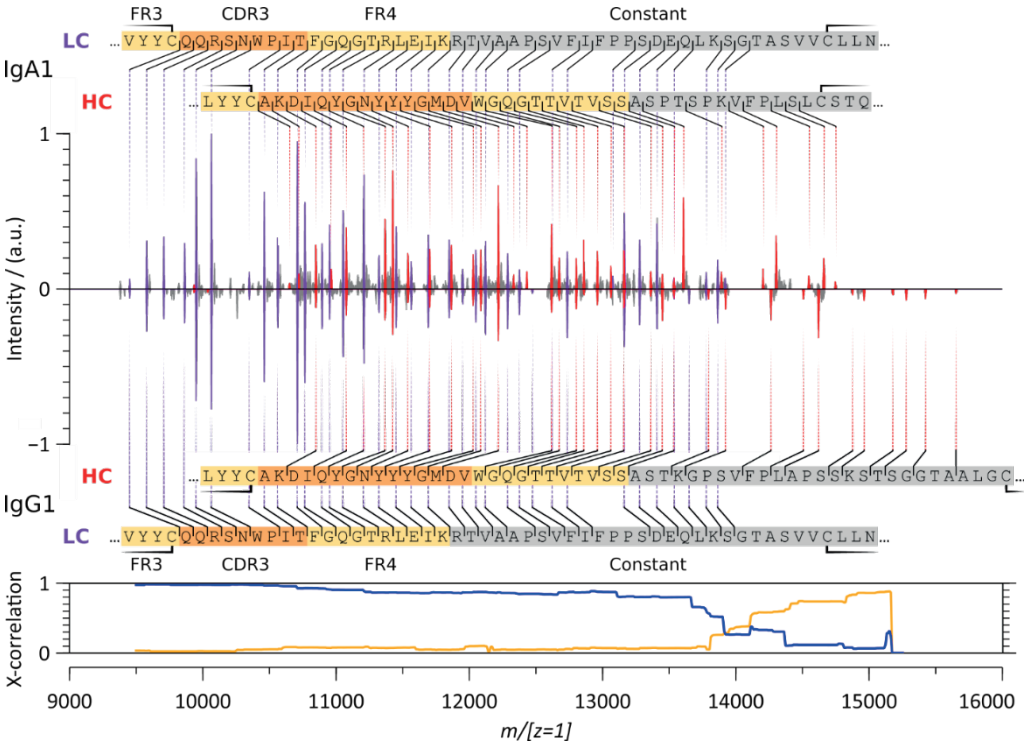


Figure 4 | Charge-deconvoluted native top-down ECD spectra of intact anti-CD20 IgA1 (top) and IgG1 (bottom). LC *c*-ion fragments are annotated in purple and the HC *c*-ion fragments in red. Cross-correlation between all fragments shows a very high similarity up to *m*/[*z* = 1] 14,000 (blue). Accounting for the known sequence deviation between the anti-CD20 IgA1 and anti-CD20 IgG1 constant regions, and the corresponding mass shift (IgA.xxxPTSPK - IgG.xxxTKGPS ≡ 40.0313 Da), restores a high cross-correlation above *m*/[*z* = 1] 14,000 (orange line). See also **Figure S2**.

To further corroborate this latter statement, we proceeded with the comparison of the ECD spectra of the intact and Fab arm molecules of the anti-CD20 IgA1 (**Figures 5** and **S3**). Middle-down analysis – here, the analysis of the Fab arms formed upon selective enzymatic digestion – offers an attractive alternative to the analysis of intact Igs as it removes the primary source of heterogeneity (the glycans located on the Igs Fcs). We observed that the native top-down ECD spectra of the 157,660 ± 170 Da extensively glycosylated anti-CD20 IgA1 and the corresponding 47,958 / 48,618 Da (depending on the OpeRATOR cleavage site, see **Tables S1** and **S2** and **Figure S4**) nonglycosylated Fab arm were nearly identical in the lower m/z region, both revealing easy-to-interpret *c*-ion fragment ladders (**Figure 5**), a similitude confirmed by cross-correlation analysis.

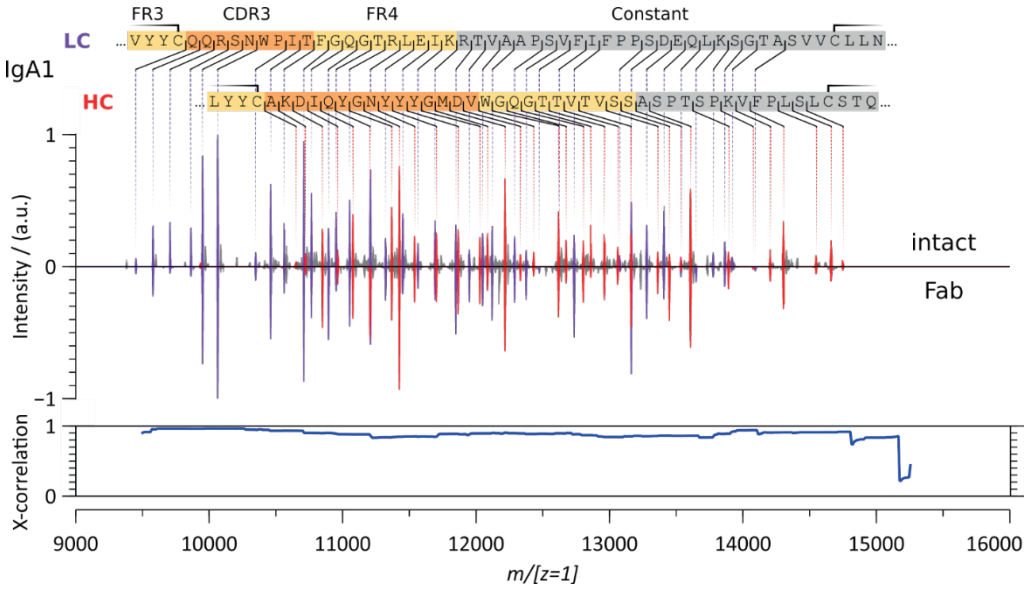


Figure 5 | Charge-deconvoluted native top-down ECD spectra of the anti-CD20 IgA1, originating from either the ~ 160 kDa highly glycosylated intact form (top) or the corresponding ~ 48 kDa nonglycosylated Fab arm (bottom). The LC *c*-ion fragments are annotated in purple and the HC fragments in red. The cross-correlation analysis (with values > 0.85 over the whole spectral range) highlights the similitude between the spectra.

An advantage of fragmenting proteins below 50 kDa in mass is that both the concomitantly formed low m/z *c*-ions and complementary high m/z *z*-ions can be isotopically resolved using high-resolution mass spectrometry (here 200,000 at m/z 400). This allows us to further corroborate the observation – first evidenced by measurements on IgG[8] – that ECD without additional vibrational or electronic excitation does not lead to substantial cleavage of disulfide bridges in native immunoglobulins (**Figure 6**).

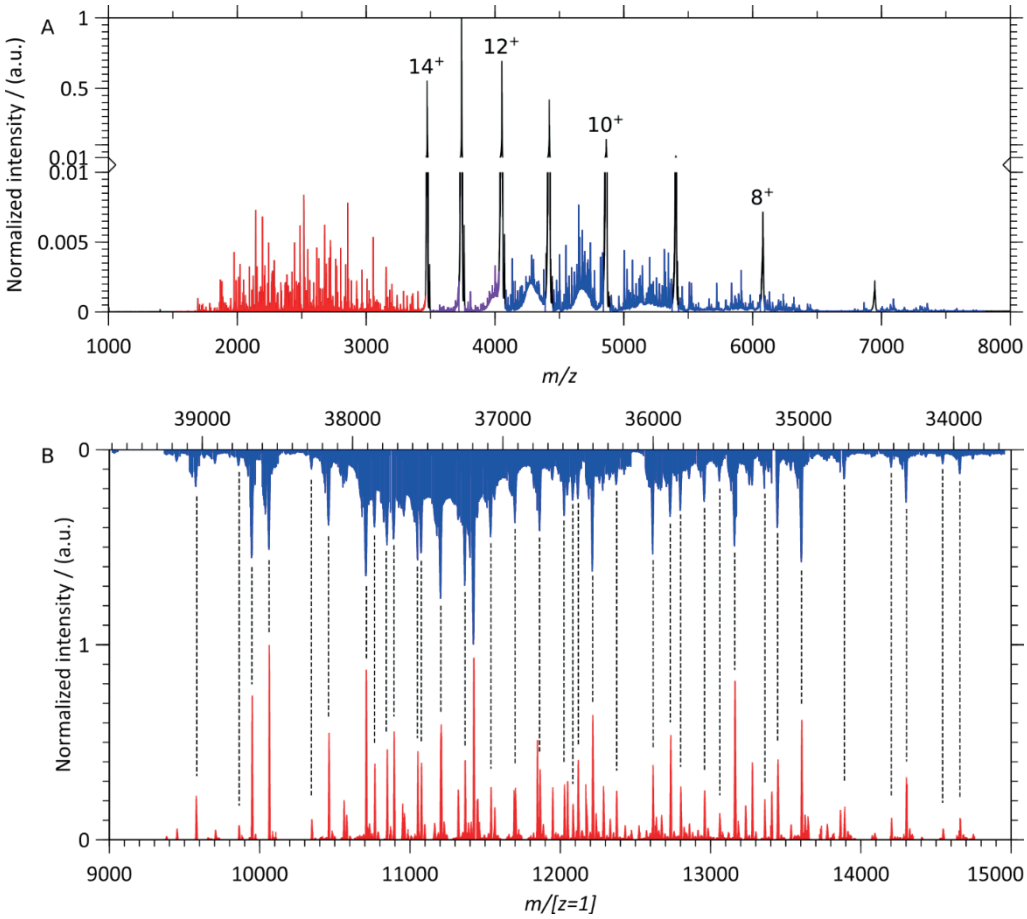


Figure 6 | Concomitant N-terminal *c*-ions and bridged *z*-ions produced in the native top-down ECD spectra of the anti-CD20 IgA1 Fab precursor, revealing that the inter- and intrachain disulfide bonds remain intact. (A) Most *c*-ions (< *m/z* 3500, red) are well separated in the *m/z* window from their concomitant bridged *z*-ions (> *m/z* 3500, blue) which overlap with the charged reduced precursors in the native ECD-MS spectrum of the IgA1 anti-CD20 Fab molecule. The purple regions correspond to the spectral region with overlapping *c*-ions and bridged *z*-ions. **(B)** Alignment of the *c*- and *z*-ion pairs, in red and blue, respectively, after charge deconvolution, reveals that their summed-up masses precisely add up to that of the intact 48,618 Da Fab precursor for each detected (*c*, bridged *z*) ion pair.

With the C-terminal cysteine of the LC forming a disulfide bridge with a cysteine on the HC, two pairs of fragments are detected for each Fab arm molecule: an LC or HC N-terminal *c*-ion paired with the complementary C-terminal disulfide-bridged *z*-ion. In contrast to intact Igs, the higher mass disulfide-bridged *z*-ion – termed "bridged" *z*-ion to highlight the conservation of the disulfide bridges between the LC and HC – can be isotopically resolved and thereby unambiguously paired with its complementary N-terminal *c*-ion, as illustrated in **Figure 6B**. While the near-perfect complementarity of the *c*- and bridged *z*-ion spectral regions confirms the conservation of the disulfide bond bridging the LC and HC, it also enables the unambiguous localization of (post-translational) modifications

affecting the CDR3 and FR4 regions or the absence thereof. As detailed in **Tables S1** and **S2** and **Figure S4**, proteolytic digestion of the anti-CD20 IgA1 results in two Fab species, one at 47,958 Da (matching the predicted cleavage site) and another at 48,618 Da (the consequence of a proteolytic miscleavage by OpeRATOR resulting in an additional TPP at the HC's C-terminus and an HexHexNAc attached). While identical *c*-ion ladders, resulting from ECD, confirm the sequence identity of both compounds, matching bridged *z*-ions unambiguously rule out the presence of labile modifications over the same sequence range.

In **Figure 7**, we compare the *c*-ion intensities of the heavily and heterogeneously *N*- and *O*-glycosylated anti-CD20 IgA1, the corresponding *N*-glycosylated anti-CD20 IgG1, and their Fab counterparts. Owing to the complete sequence identity between the LCs of the anti-CD20 IgA1 and IgG1, LC *c*-ion ladders are nearly identical. For the HC *c*-ion ladders, they are highly similar up to residue number 135, a consequence of IgA1 and IgG1 disulfide bridging patterns. Interacting LC and HC chains are organized into Fab arms stabilized by a disulfide bond between the LC C-terminal cysteine and an HC cysteine. This disulfide bond is located significantly more N-terminal in IgA1s than in IgG1s. Compared to IgG1s, the HC of IgA1 also carries two additional intrachain disulfides, one stabilizing the C-terminal end of the Fab part and the other the N-terminal area of the Fc part (**Figure 1**)[3]. Consequently, IgA1 disulfide bonds constrain fragmentation to the region between the disulfide-stabilized Ig folds, which is smaller than its IgG1 counterpart and more comparable to the pattern observed for the IgG2, IgG3, and IgG4 subclasses[8].

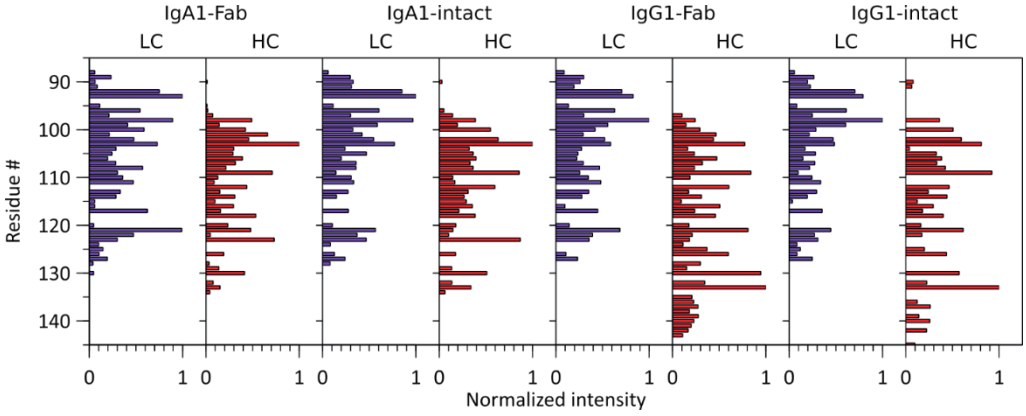


Figure 7 | Bar chart displaying normalized *c*-ion fragment intensities for each residue in the LCs and HCs of IgA1 Fab, IgA1 intact, IgG1 Fab, and IgG1 intact. See **Figure S4** for the detected and annotated much lower abundant *a*-ions.

Correlation between Fragment Ion Intensity and the Energy Required to Separate ECD Fragments

In native top-down mass spectrometry, ECD fragment ion intensities correlate with structural features ranging from protonation sites directing electron capture, and hydrogen-bonding patterns governing radical hydrogen transfer, to covalent and noncovalent interactions hindering fragment separation[7]. Here, we investigate the

correlation between fragment ion intensity and fragment stabilization by noncovalent interactions.

Briefly, as in other immunoglobulins, native IgA1 LCs and HCs, respectively, fold into two or four globular secondary structures known as the Ig folds. For both LC and HC, one of these structures is part of the variable region, while others occur in the constant regions. Typically containing about 110 amino acids, each Ig fold comprises two stacked β-sheets made up of antiparallel β-strands sandwiching a stabilizing disulfide bond[3]. Additional stabilization of polypeptide pairs relies on an array of noncovalent interactions, chiefly hydrogen bonds and van der Waals interactions, which all may influence the observed fragmentation.

In **Figure 8** and **Figure S5**, we compare ECD *c*-ion intensities to predictions corresponding to 1 plus the normalized stabilization energy of a given (*c*-, bridged *z*-) ion pair, which is negative. As expected, the fragments with the less negative stabilization energies–thus requiring the lowest activation energy to undergo separation–are also among the most abundant in the ECD spectra of Igs. While the current model describes trends reasonably well, an accurate prediction of ion intensities will have to take proton localization and hydrogen interactions with the protein backbone explicitly into account. Still, our simulations show that to some extent the ECD fragment ion spectra of immunoglobulins can be predicted.

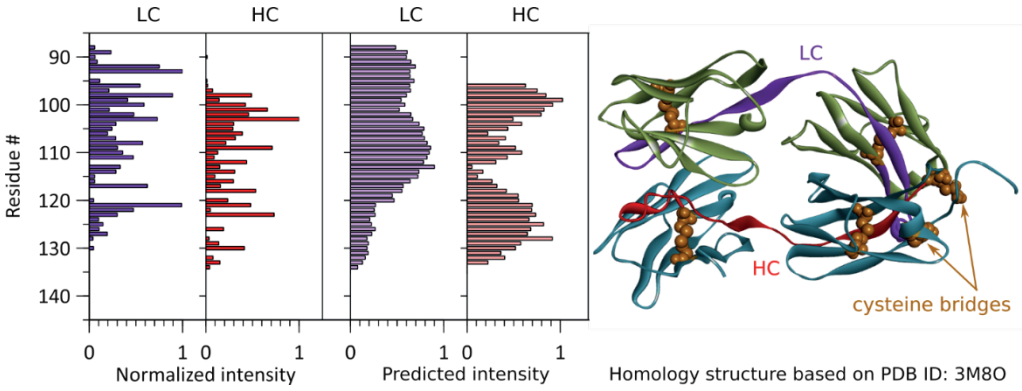


Figure 8 | Correlation between observed fragment ion intensities and intensities predicted from the normalized interaction energies of (*c*-, bridged *z*-) ion pairs for the anti-CD20 IgA1 Fab molecule using the depicted homology model, based on the crystal structure (PDB ID 3M8O) used to compute the interaction energies.

Conclusion

We explored the benefits of native top-down electron capture dissociation for the mass-spectrometry-based sequencing of immunoglobulins. Extending beyond earlier related work[8,12,32,33], we targeted a very complex heavily glycosylated recombinant immunoglobulin IgA1. Due to this structural heterogeneity, even high-resolution native mass spectrometry is not able to dissect the proteoform composition of the molecule. Overcoming this enormous structural diversity, we deliberately used native ECD as we hypothesized, based on earlier work[8,12], that this could provide sequence coverage restricted to the highly informative hypervariable CDR3 regions of both the light and heavy chains, which carry none of the reported glycosylation sites responsible for this heterogeneity.

We conducted native top-down ECD at minimal collisional activation and low pressure. We obtained straightforward-to-read amino acid sequence ladders restricted solely to *c*-ions and spanning the CDR3s and FR4s. Very little difference was observed, both qualitatively and quantitatively, between the native top-down ECD spectra of the heterogeneously *N*- and *O*-glycosylated anti-CD20 IgA1, the corresponding *N*-glycosylated anti-CD20 IgG1, and their Fab parts. The data therefore confirm our observation, initially made for the four IgG subclasses, that "pure" ECD, *i.e.*, ECD in the absence of supplemental collisional activation, results in the formation of straightforward-to-read *c*-ion ladders covering the CDR3 and FR4 of immunoglobulins independently of their class and subclass and can thus also be used to target highly glycosylated immunoglobulins. The observed ECD fragmentation spectra also can, to a large extent, be predicted quantitatively, primarily based on knowledge about the disulfide linkages between the light and heavy chains and interactions between the different domains in the immunoglobulins. Further developments, accounting for protonation sites and hydrogen bonds targeting backbone atoms, are therefore expected to open the way to spectral prediction via libraries incorporating intensity information.

Through this work, we expanded our understanding of the very specific ECD fragmentation occurring in the CDR3 and FR4 regions of different classes of immunoglobulins. We conclude that the optimized ECD method and spectrum prediction algorithms can become very powerful and beneficial for the de novo sequencing of human immunoglobulins.

Notes

The authors declare the following competing financial interest(s): F.J.B. and J.S. are Genmab employees and have stock and/or warrants. This work was also partly supported by Genmab through funding for A.B., S.T., and A.J.R.H.

Acknowledgements

We thank the members of the Heck laboratory for general support, especially Arjan Barendregt. J.F.G. thanks Joshua Klein (Boston University CBMS) for modifications to the MSDeisotope python library. This research received funding through The Netherlands Organization for Scientific Research (NWO) ENPPS.LIFT.019.001 project (A.J.R.H. and J.F.G.), the NACTAR project 16442 (A.J.R.H. and M.A.d.B.), and the Spinoza Award SPI.2017.028 to A.J.R.H.

References

1. Orange, J.S. et al. Use of intravenous immunoglobulin in human disease: A review of evidence by members of the Primary Immunodeficiency Committee of the American Academy of Allergy, Asthma and Immunology. *Journal of Allergy and Clinical Immunology* **117**, S525-S553 (2006).
2. Leusen, J.H.W. IgA as therapeutic antibody. *Mol. Immunol.* **68**, 35-39 (2015).
3. de Sousa-Pereira, P. & Woof, J.M. IgA: Structure, Function, and Developability. *Antibodies* **8**, 57 (2019).
4. de Haan, N., Falck, D. & Wuhrer, M. Monitoring of immunoglobulin N- and O-glycosylation in health and disease. *Glycobiology* **30**, 226-240 (2020).
5. Zubarev, R.A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **15**, 12-16 (2004).
6. Taouatas, N., Drugan, M.M., Heck, A.J.R. & Mohammed, S. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* **5**, 405-407 (2008).
7. Lermyte, F., Valkenborg, D., Loo, J.A. & Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* **37**, 750-771 (2018).
8. den Boer, M.A., Greisch, J.-F., Tamara, S., Bondt, A. & Heck, A.J.R. Selectivity over coverage in de novo sequencing of IgGs. *Chemical Science* **11**, 11886-11896 (2020).
9. Mao, Y., Valeja, S.G., Rouse, J.C., Hendrickson, C.L. & Marshall, A.G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **85**, 4239-4246 (2013).
10. Shaw, J.B. et al. Sequencing Grade Tandem Mass Spectrometry for Top–Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Anal. Chem.* **90**, 10819-10827 (2018).
11. Ganisl, B. & Breuker, K. Does Electron Capture Dissociation Cleave Protein Disulfide Bonds? *ChemistryOpen* **1**, 260-268 (2012).
12. Shaw, J.B. et al. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **92**, 766-773 (2020).
13. Teeling, J.L. et al. Characterization of new human CD20 monoclonal antibodies with potent cytolytic activity against non-Hodgkin lymphomas. *Blood* **104**, 1793-1800 (2004).
14. Bondt, A. et al. Immunoglobulin G (IgG) Fab Glycosylation Analysis Using a New Mass Spectrometric High-throughput Profiling Method Reveals Pregnancy-associated Changes*. *Mol. Cell. Proteomics* **13**, 3029-3039 (2014).
15. Fort, K.L. et al. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **17**, 926-933 (2018).
16. Pacholarz, K.J. et al. Molecular Insights into the Thermal Stability of mAbs with Variable-Temperature Ion-Mobility Mass Spectrometry. *ChemBioChem* **17**, 46-51 (2016).
17. Chambers, M.C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918-920 (2012).
18. Klein, J., Carvalho, L. & Zaia, J. Application of network smoothing to glycan LC-MS profiling. *Bioinformatics* **34**, 3511-3518 (2018).
19. Lukauskas, J.K.h.S. (Zenodo, 2019).
20. Gibbons, C.W.B. (Computational Mass Spectrometry @ Pacific Northwest National Laboratory, 2020).

21. Park, J. et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **14**, 909-914 (2017).
22. Venable, J.D., Xu, T., Cociorva, D. & Yates, J.R. Cross-Correlation Algorithm for Calculation of Peptide Molecular Weight from Tandem Mass Spectra. *Anal. Chem.* **78**, 1921-1929 (2006).
23. Fiser, A. & Šali, A. in Methods in Enzymology, Vol. 374 461-491 (Academic Press, 2003).
24. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382-W388 (2005).
25. Schymkowitz, J.W.H. et al. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci.* **102**, 10147-10152 (2005).
26. Tomana, M., Niedermeier, W., Mestecky, J. & Skvaril, F. The differences in carbohydrate composition between the subclasses of IgA immunoglobulins. *Immunochemistry* **13**, 325-328 (1976).
27. Maurer, M.A. et al. Glycosylation of Human IgA Directly Inhibits Influenza A and Other Sialic-Acid-Binding Viruses. *Cell Rep.* **23**, 90-99 (2018).
28. Field, M.C., Amatayakul-Chantler, S., Rademacher, T.W., Rudd, P.M. & Dwek, R.A. Structural analysis of the N-glycans from human immunoglobulin A1: comparison of normal human serum immunoglobulin A1 with that isolated from patients with rheumatoid arthritis. *Biochemical Journal* **299**, 261-275 (1994).
29. Mattu, T.S. et al. The Glycosylation and Structure of Human Serum IgA1, Fab, and Fc Regions and the Role of N-Glycosylation on Fcα Receptor Interactions*. *J. Biol. Chem.* **273**, 2260-2272 (1998).
30. Royle, L. et al. Secretory IgA N- and O-Glycans Provide a Link between the Innate and Adaptive Immune Systems*. *J. Biol. Chem.* **278**, 20140-20153 (2003).
31. Rosati, S. et al. Exploring an Orbitrap Analyzer for the Characterization of Intact Antibodies by Native Mass Spectrometry. *Angew. Chem. Int. Ed.* **51**, 12992-12996 (2012).
32. Fornelli, L. et al. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **90**, 8421-8429 (2018).
33. He, L. et al. Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS. *J. Am. Soc. Mass. Spectrom.* **28**, 827-838 (2017).

Supporting Information

Supplementary Figures and Tables

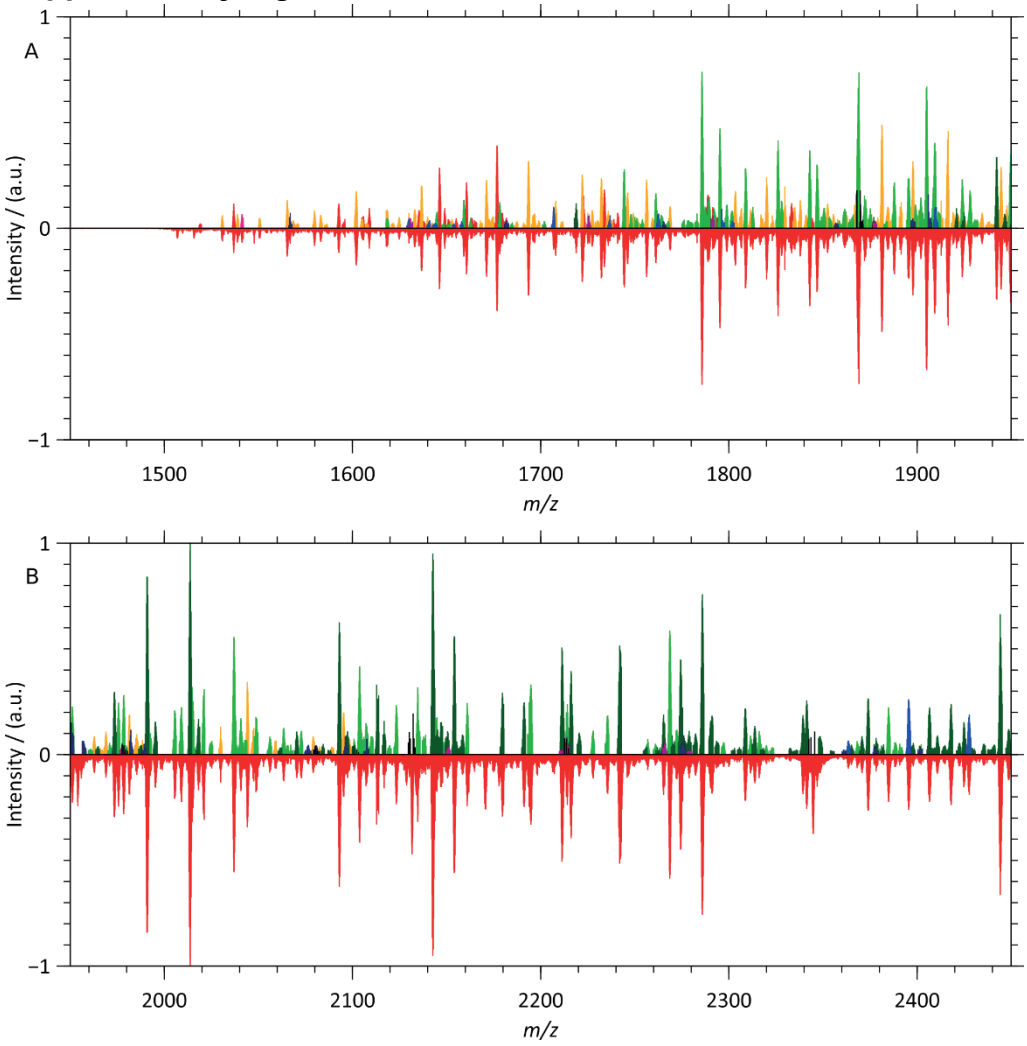


Figure S1A, B | Comparison of the low *m/z* region of the native top-down ECD spectra of intact anti-CD20 IgA1 spectrum reconstructed from charge-deconvoluted spectra corresponding to fragment charge states ranging from 1+ to 13+ (top), with the unprocessed fragment ion spectrum (bottom). (A) [1450, 1950] *m/z* spectral range. **(B)** [1950, 2450] *m/z* spectral range. See **Figure 3** for the complete spectral range.

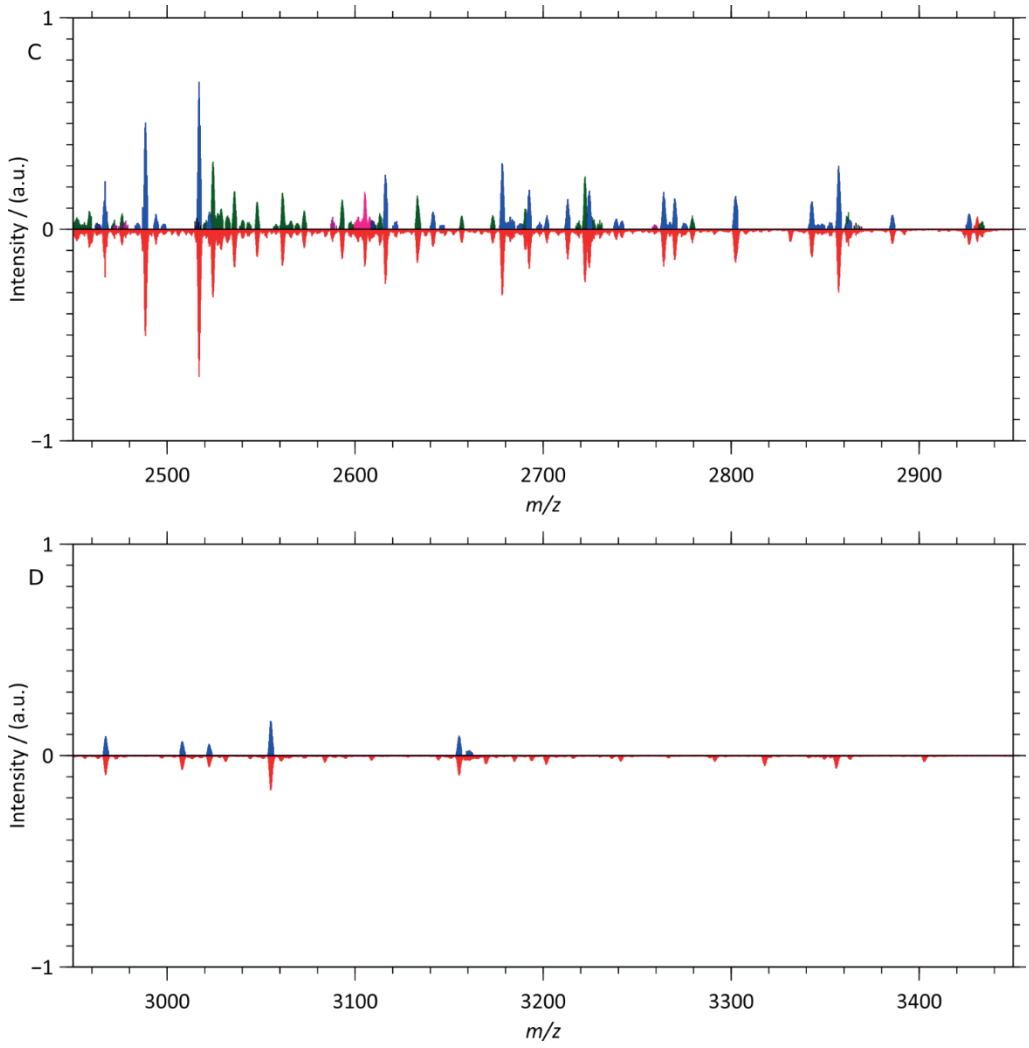


Figure S1C, D | Comparison of the low *m/z* region of the native top-down ECD spectra of intact anti-CD20 IgA1 spectrum reconstructed from charge-deconvoluted spectra corresponding to fragment charge states ranging from 1+ to 13+ (top), with the unprocessed fragment ion spectrum (bottom). (C) [2450, 2950] *m/z* spectral range. **(D)** [2950, 3450] *m/z* spectral range. See **Figure 3** for the complete spectral range.

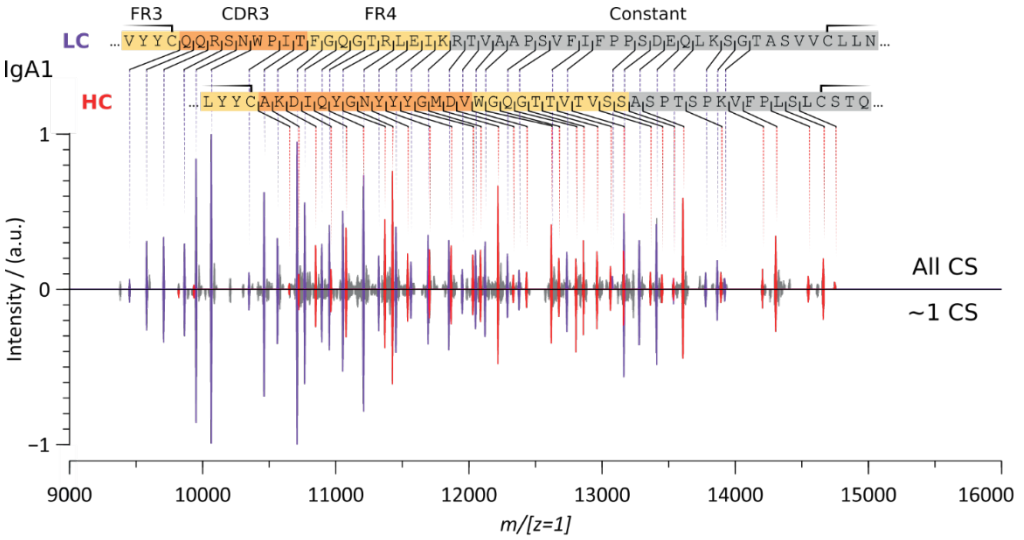


Figure S2 | Charge-deconvoluted native top-down ECD spectra of intact anti-CD20 IgA1 selecting as precursor either all charge states (top) or a much narrower isolation window of 210 *m/z* targeting the charge state 25+ (bottom). LC *c*-ion fragments are annotated in purple and the HC *c*-ion fragments in red. Notably, the resulting ECD spectra look nearly identical.

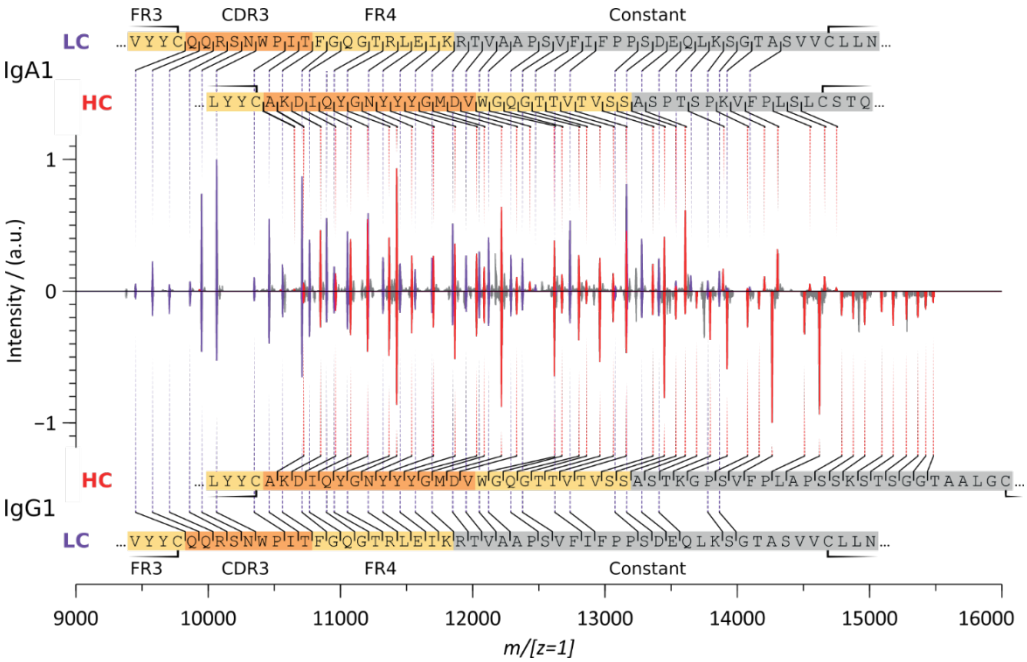


Figure S3 | Charge-deconvoluted native top-down ECD spectra of anti-CD20 IgA1's (top) and IgG1's (bottom) Fab arms. LC *c*-ion fragments are annotated in purple and the HC *c*-ion fragments in red.

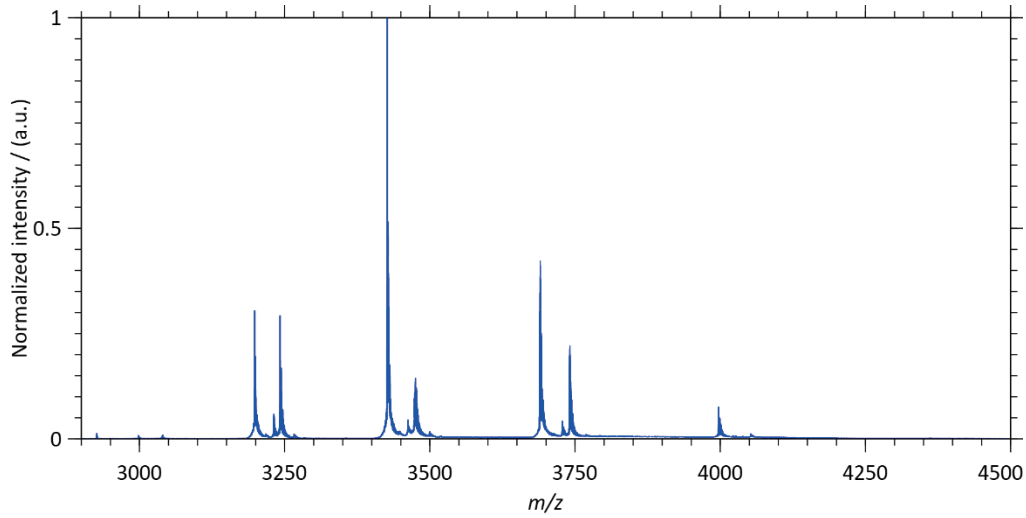


Figure S4 | MS1 spectrum of the anti-CD20 IgA1 Fab. Two distinct Fab fragments are observed of 47,958 Da and 48,618 Da with the latter due to a missed cleavage of the OpeRATOR enzyme, adding 295 Da (TPP) and 365 Da (HexNAc + Hex) to the mass of the main fragment.

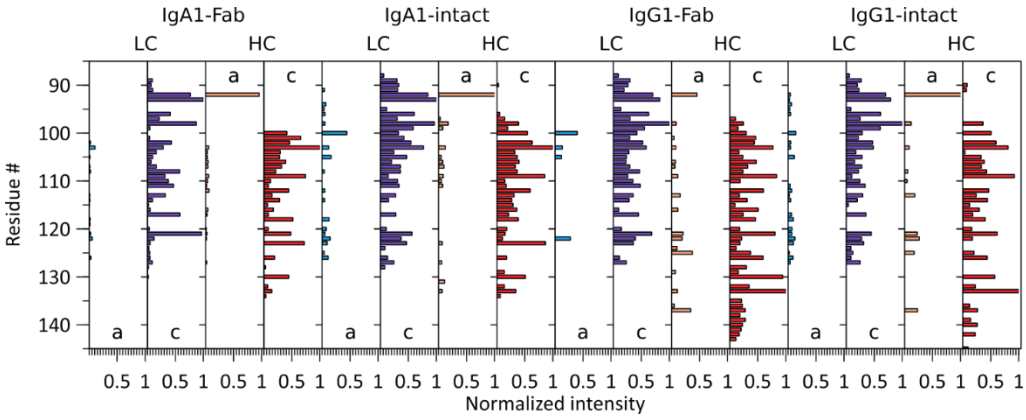


Figure S5 | Bar-chart displaying normalized *c*- and *a*-ion fragment intensities for each residue in the LCs and HCs of IgA1 Fab, IgA1 intact, IgG1 Fab, and IgG1 intact. The *a*-ion fragments are generally of lower abundance, but for some cleavage sites, a-ions complement missing *c*-ions.

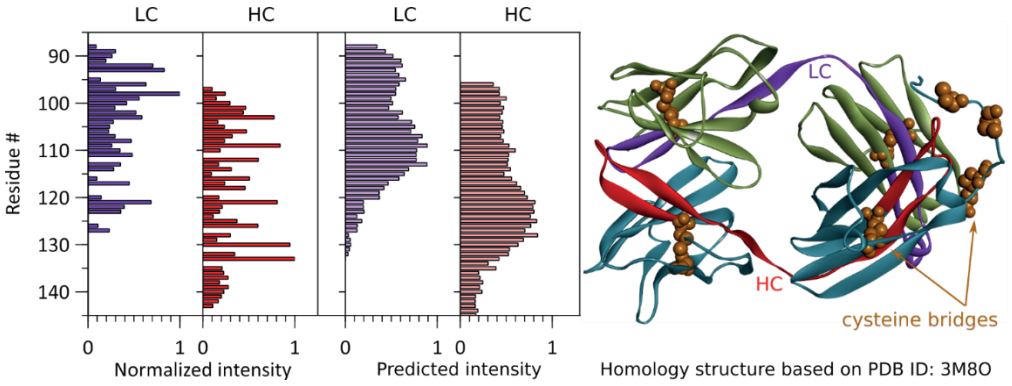


Figure S6 | Correlation between fragment ion intensities and intensities predicted from the normalized interaction energies of (*c*-, bridged *z*-) ion pairs for the anti-CD20 IgG1 Fab molecule and the homology model, based on the available crystal structure (PDB ID 3BKY), used to compute the interaction energies.

Table S1 | Amino acid sequences of the anti-CD20 IgG1 and IgA1 mAbs used. The primary cleavage sites of the IgdE (IgG1) and OpeRATOR (IgA1) enzymes that produce the Fab molecules are indicated with a bold forward slash in the heavy chain sequences.

IgG1 anti-CD20 (7D8)

Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRFTISRDNAKNSLYLQMNSLR
AEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHT
FPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKT/HTCPPCPAPELLGGPSVFLFPPKPKDTLMISR
TPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQ
PREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMH
EALHNHYTQKSLSLSPGK

κ Light Chain

EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLTISSLEPEDFAVYY
CQQRSNWPITFGQGTRLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYS
LSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

IgA1 anti-CD20 (7D8)

Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRFTISRDNAKNSLYLQMNSLR
AEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASPTSPKVFPLSLCSTQPDGNVVIACLVQGFFPQEPLSVTWSESGQGVTARN
FPPSQDASGDLYTTSSQLTLPATQCLAGKSVTCHVKHYTNPSQDVTVPCPVPS/TPP/TPSPSTPPTPSPSCCHPRLSLHRPALEDL
LLGSEANLTCTLTGLRDASGVTFTWTPSSGKSAVQGPPERDLCGCYSVSSVLPGCAEPWNHGKTFTCTAAYPESKTPLTATLSKSGN
TFRPEVHLLPPPSEELALNELVTLTCLARGFSPKDVLVRWLQGSQELPREKYLTWASRQEPSQGTTTFAVTSILRVAAEDWKKGDTF
SCMVGHEALPLAFTQKTIDRLAGKPTHVNVSVVMAEVDGTCY

κ Light Chain

EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLTISSLEPEDFAVYY
CQQRSNWPITFGQGTRLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYS
LSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Table S2 | Overview of the average molecular masses of the intact Igs and Fab molecules used in this study. The theoretical mass is obtained from the provided protein sequences as shown in **Table S1**, the experimental mass is determined by native mass spectrometry.

Species	$MW_{theoretical}$ (Da)	$MW_{experimental}$ (Da)	Mass difference (Da)
IgG1 anti-CD20	146,228	148,858	2,630*
IgA1 anti-CD20	149,191	(avg. gly.) 157,660	8,469**
IgG1 anti-CD20 Fab	47,900	47,900	-0.8
IgA1 anti-CD20 Fab	47,956.6	47,957.5	0.9
IgA1 anti-CD20 Fab	48,618.1***	48,617.9	-0.2

* Corresponding to the mass shift induced by G0F/G0F glycosylation and *C*-terminal lysine clipping of both HCs.
** Induced by extensive *N*- and *O*-glycosylation of the IgA1 HCs.
*** Due to an additional TPP at the HC's *C*-terminus and an HexHexNAc attached (at T).

Table S3 | ECD cell parameters.

Species	L1	L2	LM3	L4	FB	LM5	L6	L7
Immunoglobulins	2.12	-13.2	7.0	4.7	-1.5	7.0	-15.2	1.5

CHAPTER 9

EXTENDING NATIVE TOP-DOWN ELECTRON CAPTURE DISSOCIATION TO MDA IMMUNOGLOBULIN COMPLEXES PROVIDES USEFUL SEQUENCE TAGS COVERING THEIR CRITICAL VARIABLE COMPLEMENTARITY-DETERMINING REGIONS

Jean-Francois Greisch[1,2], Maurits A. den Boer[1,2], Szu-Hsueh Lai[1,2], Kelly Gallagher[1,2], Albert Bondt[1,2], Jan Commandeur[3], and Albert J.R. Heck[1,2]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

[2] *Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands*

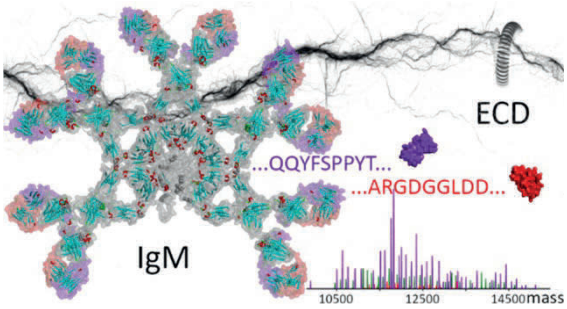
[3] *MSVision, Televisieweg 40, 1322 AM Almere, The Netherlands*

Published as:

Extending Native Top-Down Electron Capture Dissociation to MDa Immunoglobulin Complexes Provides Useful Sequence Tags Covering Their Critical Variable Complementarity-Determining Regions
Jean-Francois Greisch, Maurits A. den Boer, Szu-Hsueh Lai, Kelly Gallagher, Albert Bondt, Jan Commandeur, and Albert J.R. Heck
Analytical Chemistry 2021, 93, 48, 16068-16075
DOI: 10.1021/acs.analchem.1c03740

Abstract

Native top-down mass spectrometry (MS) is gaining traction for the analysis and sequencing of intact proteins and protein assemblies, giving access to their mass and composition, as well as sequence information useful for identification. Herein, we extend and apply native top-down MS, using electron capture dissociation, to two submillion Da IgM- and IgG-based oligomeric immunoglobulins. Despite structural similarities, these two systems are quite different. The ~ 895 kDa noncovalent IgG hexamer consists of six IgG subunits hexamerizing in solution due to three specifically engineered mutations in the Fc region, whereas the ~ 935 kDa IgM oligomer results from the covalent assembly of one joining (J) chain and 5 IgM subunits into an asymmetric "pentamer" stabilized by interchain disulfide bridges. Notwithstanding their size, structural differences, and complexity, we observe that their top-down electron capture dissociation spectra are quite similar and straightforward to interpret, specifically providing informative sequence tags covering the highly variable CDR3s and FR4s of the Ig subunits they contain. Moreover, we show that the electron capture dissociation fragmentation spectra of immunoglobulin oligomers are essentially identical to those obtained for their respective monomers. Demonstrated for recombinantly produced systems, the approach described here opens up new prospects for the characterization and identification of IgM clones circulating in plasma, which is important since IgM plays a critical role in the early immune response to pathogens such as viruses and bacteria.



Introduction

Early control of viral and bacterial infections is dependent on innate natural antibodies. Among those, immunoglobulins M (IgM) are critical to the initial humoral immune response[1-3]. Via their ability to recognize modified self-components and altered cells displaying specific patterns such as carbohydrates, glycolipids, and repetitive structures, IgMs also contribute to immunosurveillance mechanisms against precancerous and cancerous cells[4,5]. Consequently, IgMs are promising agents for immunotherapy[4,6].

IgM's high molecular weight, high number of isoforms, and rather complex structure pose, however, a challenge to both therapeutic and diagnostic applications. The vast majority of IgMs consists of five identical subunits and one joining (J) chain assembled into an asymmetric pentamer[7-9]. Besides being present in the gastrointestinal tract, lymphatic vessels, mucosal tissues, bone marrow, and so forth, IgM also constitutes about 30% of the blood-circulating immunoglobulins[10]. Similarly to other isotypes, IgM's subunits are formed by two heavy chains (HCs), each paired to one light chain (LC) and stabilized by intra- and interchain disulfide bonds[11]. Stabilization of the pentameric assembly relies on interactions of the C-terminal of the HCs involving inter-subunit disulfide bonds, the formation by the C-terminal 18 amino-acid-long secretory tailpieces of a central β-sandwich structure, and bonding to the J-chain[8,9,12]. The mature J-chain contains eight cysteine residues with two involved in disulfide bonds with an IgM HC and the other six forming intrachain disulfide bridges[13].

IgM naturally forms oligomers, whereas IgGs are thought to predominantly exist as monomers. The introduction of specific mutations in recombinant IgG1s can, however, induce the formation of stable IgG hexamers in solution[14-17]. Compared to disulfide-stabilized IgMs, these hexameric IgG1s are characterized by more mobile Fab regions – reduced steric hindrance related to the long IgG hinge – which may facilitate binding to low-accessibility or close-to-the-membrane epitopes[14]. The therapeutic potential of these induced IgG hexamers is currently being investigated, especially for their role in complement activation[17-20].

Here, we aim at analyzing these very large and important, albeit structurally complicated, oligomeric immunoglobulins, through the application of native top-down mass spectrometry (nTDMS)[21-24]. Relying on the fact that native MS transfers the molecules to the gas phase under conditions retaining their structural features, native top-down proteomics should provide access to the sequence of the proteins' subunits and an insight into its post-translational modifications and higher-order structure. It has become apparent that inducing fragmentation in native top-down proteomics requires moving beyond the well-established activation method of collision-induced dissociation (CID) as this rather slow heating process often induces too little fragmentation when applied to these very high mass systems[25-27]. Several groups have explored alternative fragmentation methods for nTDMS, such as surface-induced dissociation[28], photon-induced dissociation (UVPD and IRMPD)[29-36], electron transfer dissociation[36,37], and electron capture dissociation (ECD)[22,34,35,38-46].

Here, we discuss the performance and characteristics of ECD (comparing it with CID) in the native top-down characterization of two ~ 1 MDa immunoglobulin oligomers, making use of (1) a recombinant IgM pentamer with J-chain targeting the wall teichoic acids (WTA) antigen (molecular weight of approximately 937,500 Da) and (2) an engineered recombinant IgG1-RGY hexamer (molecular weight of approximately 895,300 Da) targeting CD52.

We *a priori* hypothesized that the native top-down analysis of these two systems would be analytically very challenging or even impossible, for various reasons. First, their size and high mass (close to 1 MDa) make them challenging to ionize and analyze by MS. Second, their vast structural complexity – especially for the $(IgM)_5$:$(J)_1$ pentamer with J-chain – resulting from disulfide bonds and glycans, leads to heterogeneous mass distributions and complicates fragment ion formation. Third, being built of roughly 7000 amino acids, extensive backbone cleavage by nTDMS may theoretically result in overwhelmingly complex fragmentation spectra.

We showed in previous work that ECD without supplemental activation of monomeric IgG (~ 150 kDa) and IgA (~ 165 kDa) variants can result in very clean and interpretable fragmentation spectra dominated by ($c/z\cdot$) fragment pairs without side-chain cleavages[45,46]. Targeting a single type of bond for fragmentation reduces spectral congestion and the dilution of signal intensity in the (n)TDMS spectra of native intact antibodies[47,48]. The conservation of inter- and intrachain disulfide bridges under ECD conditions also results in straightforward-to-read *c*-ion sequence ladders from sequence segments outside regions bridged by disulfide-bridged loops. For the system studied, the fragments cover the LC and HC CDR3s, as well as the sequence segments, immediately downstream without interference of fragments from other parts of the protein. Testing the ECD nTDMS approach on all IgG subtypes (IgG1, IgG2, IgG3, and IgG4) and an IgA1 proved the method to be generally applicable to monomeric immunoglobulins even when they are heavily glycosylated[47,49].

As an ambitious follow-up, we here explore the application of ECD without supplemental activation to the 5-6 times larger covalently linked J-chain-coupled IgM pentamers and the noncovalently associated IgG1 hexamers. ECD without supplemental activation of these ~ 900-950 kDa assemblies primarily leads to extensive electron capture without dissociation (ECnoD) with sometimes up to 40 electrons captured by the precursor ions. However, in the low m/z range of both the J-chain-coupled IgM pentamers and IgG1-RGY hexamers, informative fragment ions are observed, which yield sequence information for the complementarity-determining regions (CDRs), similarly to the monomers. When comparing ECD and CID fragmentation behaviors for these systems, we observe that informative sequence tags from the variable regions of oligomeric IgG and IgM are favorably generated by native top-down ECD. The different nature of these two immunoglobulin oligomers, one, a highly glycosylated covalently linked heterohexamer and the other a noncovalent homohexamer, does not seem to significantly impact the formation of the CDR sequence-informative low m/z fragments. As in the case of refs[45,46], ECD appears here directed by charge localization in the vicinity of the *N*-terminal regions. Consequently, the present approach suggests that serum-purified IgM molecules could

also, in principle, be partly *de novo* sequenced by this native top-down approach, thereby facilitating the analysis of this important class of antibodies.

Materials and Methods

Samples

The monoclonal aWTA $(IgM)_5$:$(J)_1$ was provided by Suzan Rooijakers (Medical Microbiology, UMCU, The Netherlands). The aCD52 hexamer-forming IgG1-RGY mutant[14,16] was provided by Genmab (Utrecht, The Netherlands).

Preparation of Intact $(IgM)_5$:$(J)_1$ and RGY-IgG1 Hexamer Samples for Native Top-Down MS Analysis

Prior to the native top-down analysis, the storage buffer was exchanged to aqueous 150 mM ammonium acetate (pH ~ 7.0) through six consecutive dilutions and concentration steps at 4 °C using Amicon Ultra centrifugal filters with a 10 kDa molecular weight cutoff (Merck KGaA, Darmstadt, Germany). $(IgM)_5$:$(J)_1$ and hexameric IgG-RGY concentrations used for native electrospray ionization MS were typically around 2 µM (concentration of the monomer).

Native Top-Down ECD and CID MS

Top-down MS of native immunoglobulin oligomers (and monomers) was performed on a Q Exactive ultrahigh mass-range (UHMR) Orbitrap (Thermo Fisher Scientific, Bremen, Germany) using a built-in collision cell and an ECD cell developed by e-MSion[50]. Electrospray involved a nanospray source using emitters produced from borosilicate capillaries by a P-97 Sutter puller (Novato, CA, USA). ESI voltages were in the 1.0-1.4 kV range, the source temperature was set at 275 °C, and the S-lens RF was set at 200 (service mode). The in-source trapping desolvation voltage was set to -100 V as it corresponds to an optimum in terms of fragmentation[51]. Using ion transfer parameters optimized for high m/z ions, the $(IgM)_5$:$(J)_1$ or $(IgG)_6$ intact precursor ions were guided to the quadrupole, where they underwent mass selection using windows narrowed to the intact precursor charge distribution (see **Table S1**). The ions were then transferred into the ECD cell, where they were transmitted in the absence of electrons to the HCD cell for CID or subjected to electron capture dissociation for ECD. The small permanent magnets responsible for the radial confinement of electrons emitted by a cathode and the additional electrodes responsible for the electron cloud longitudinal confinement ensured efficient ion transfer to the HCD cell in both cases. Upon transfer from the ECD cell to the HCD cell following electron capture, potential post-ECD collisional activation was kept to a minimum (HCD Direct eV setting = 1 or 0) to avoid (b/y) ion formation and to concentrate the fragment signal intensity into the *c*- and *z*-ions of interest, while for CID, an HCD direct eV setting = 200 was used (**Figure S1** and **Table S1**). The analyzer injection parameter, CE-Inject (V) UHMR, was initially kept to its standard value of 3200 for high m/z ions. This ensured optimal detection of the high m/z precursor ions but proved detrimental to the recording of fragments in the 1000-3000 m/z range as the third harmonic fraction of the precursor ion signal then overlapped with the fragment ions. To reduce interference via the optimization of trajectories in the analyzer for low m/z ions, we set the CE-Inject (V)

UHMR parameter to 3700, which considerably increased the low m/z ion signal while suppressing the signal (third harmonic included) corresponding to the intact precursor. The 3700 value reduces the recording time and facilitates processing (*e.g.*, it removes the need for baseline subtraction), but it does not affect the number of detected fragments for high-quality spectra. It is worth noting that operating at a low pressure – the UHV readout is under the range, indicative of a pressure below 5×10^{-11} mbar close to the analyzer region of the instrument – is also beneficial to the recording of high-resolution spectra as it enables the detection of long transients with limited dephasing by collisions. Overall, the HCD cell trapping and extraction parameters were optimized for operation under low nitrogen collision gas pressure conditions. All spectra were acquired with the noise threshold parameter set to 3.64 at a set resolution of 280,000 @ m/z 200. Between 5000 and 10,000 scans were averaged for each spectrum.

Data Analysis

Processing of the ECD fragmentation spectra involved the conversion of raw files to the mzML format by Proteowizard[52]. We used the MSDeisotope Python library (Joshua Klein, Boston University CBMS)[53,54] with a minimum_score = 10.0 and mass_error_tolerance = 0.02 to generate a charge-deconvoluted spectrum with all the isotopic peaks retained[53,54]. Fragments were assigned by applying LcMsSpectator (Pacific Northwest National Laboratory)[55,56] to the charge-deconvoluted spectra generated by MSDeisotope. The accuracy threshold was set to 3 ppm for all assignments following recalibration of the fragment's m/z via shifting by the average error on assignable *c*-ion fragments. The results were exported as tsv files for further analysis. Sequence assignment accommodated the major ECD ($a\cdot$, c, x, z, $z\cdot$) or CID (b,y) ion types without considering H_2O and NH_3 neutral losses, except when explicitly mentioned.

Results and Discussion

Structures and Native Mass Spectra of the Studied IgM- and IgG-Based Oligomeric Immunoglobulin Molecules

The here-studied immunoglobulin oligomers, the $(IgM)_5$:$(J)_1$ pentamer and $(IgG1)_6$ hexamer, are quite different in structural organization, with the first one being a highly glycosylated and covalently linked heterohexamer (IgM pentamer with J-chain) and the second one being a noncovalent homohexamer, as schematically depicted in **Figure 1**. The recombinant production of the hexamerizing IgG1-RGY is reasonably straightforward and similar to the production of normal IgG1 antibodies, whereby the inclusion of the RGY mutations promotes extensive hexamerization in solution, as previously described[14,20,57]. The production of recombinant IgM, especially in its $(IgM)_5$:$(J)_1$ "pentamer" format, is less straightforward as it requires the appropriate coexpression of the joining J chain and IgM subunits and the correct formation of all inter- and intrachain disulfide bridges (**Figure 1A, B**). Here, a pure aWTA $(IgM)_5$:$(J)_1$ pentamer was recombinantly expressed and purified by SEC.

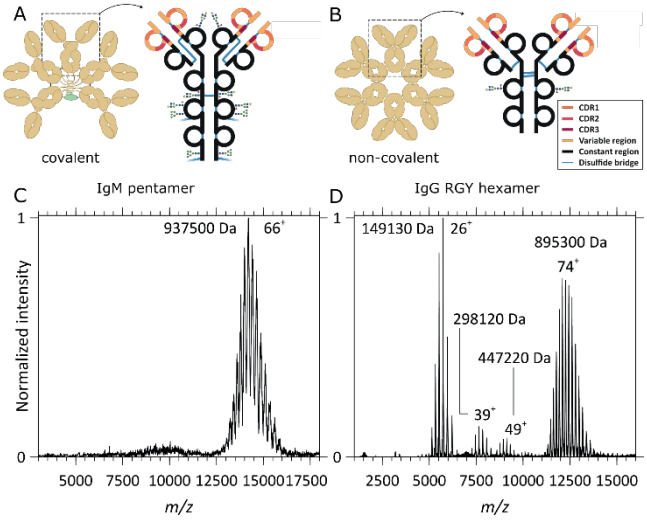


Figure 1 | Schematic structures and native MS1 spectra of oligomeric immunoglobulins. (A, C) Structure and native MS1 spectrum of the aWTA IgM pentamer with the J-chain, respectively, and **(B, D)** structure and native MS1 spectrum of the aCD52 IgG1-RGY hexamer, respectively. While, in $(IgM)_5$:$(J)_1$, the IgM monomers and J-chain are connected by interchain disulfide bonds, the IgG1 monomers forming the IgG1 hexamers assemble noncovalently. **(C)** The MS1 spectrum of the aWTA IgM displays a single charge distribution with charge state broadening resulting from the presence of multiple heterogeneous glycans on each monomer. Although heterogeneity and glycan lability hamper an accurate mass analysis, a mean average mass of 937,500 Da could be extracted from this high-resolution data set. **(D)** In the MS1 spectrum of the aCD52 IgG1-RGY, monomers, dimers, and trimers co-occur with the hexamers, a direct consequence of the noncovalent interactions stabilizing the assembly and of the dynamical equilibria taking place in solution. An average mass of 895,300 Da could be extracted from these data for the hexamer, indeed being within the experimental error 6 times that of the monomer mass.

In **Figure 1**, we depict native MS1 spectra of the aWTA IgM pentamer coupled to the J-chain and spectra of the aCD52 IgG1-RGY hexamer. For the aWTA $(IgM)_5$:$(J)_1$, a single charge-state distribution is observed in the $12{,}500 < m/z < 16{,}000$ range. Glycosylation of the aWTA IgM is quite heterogeneous, resulting in a broadening of the peaks. Although the presence of all these glycans hampers an accurate mass analysis, a mean average mass of 937,500 Da could be extracted from these data, which is in line with the theoretically expected mass of the $(IgM)_5$:$(J)_1$. The native MS spectra of the aCD52 IgG1-RGY mutant reveal charge states corresponding to IgG monomers and lower abundant dimers and trimers that co-occur with the IgG1 hexamers, a direct consequence of the noncovalent interactions stabilizing the assembly of the dynamical equilibria taking place in solution. Compared to the aWTA IgM, the aCD52 hexameric IgG1 contains a much more homogeneous glycosylation profile, as can be expected from the smaller number of glycosylation sites: IgG1 HCs harbor one *N*-linked glycosylation site in the fragment-crystallizable (Fc) region. Therefore, the ion signals in **Figure 1D** are much sharper than those in **Figure 1C**. The spectra shown in **Figure 1D** and the extracted masses are in excellent agreement with previously reported data for the IgG1 RGY constructs[20].

Native Top-Down MS of Oligomeric Immunoglobulins by CID

As displayed in **Figure 2**, we first performed CID on mass-selected aWTA $(IgM)_5$:$(J)_1$ and aCD52 hexameric IgG1 assemblies. In peptide and protein centric MS, CID is still the most ubiquitous activation technique. Applied to large noncovalent assemblies, it typically results in the release of one or more subunits carrying most of the complex charge: a process called asymmetric dissociation[58,59]. Here, we could only achieve CID by using the highest available collision energies, that is, CID of such large systems is demanding. Applied to precursor ions of aWTA $(IgM)_5$:$(J)_1$ (**Figures 2A** and **S2A**) carrying 66 charges, CID primarily results in the release of the intact IgM LC – as a result of disulfide bond cleavage. Additionally, limited backbone cleavage yields small sequence tags formed by *b*-ions covering part of the LC constant regions (in the m/z window below 5000). Contrastingly, for the precursor ions of the aCD52 hexameric IgG1 assembly which carry ~ 74 charges (**Figures 2B** and **S3A**), the primary observed CID channel is the ejection of IgG monomer ions, as expected from this noncovalent assembly. These monomer ions take, upon release, on average, one-third of all charges of the precursor ions. Secondary lower abundant dissociation channels correspond, for the hexamer, to dimer ejection and the formation of a small sequence tag made of *b*-ions covering part of the LC constant region (in the m/z window below 5000). Overall, obtaining sequence information from large (~ 900 kDa) Ig complexes using CID is challenging because of the many observed competitive dissociation channels, overlapping ion signals – the result of extensive water and ammonia losses – as well as signal broadening when recording ion signals (transients) at a resolution of 280,000 @m/z 200 and a pressure setting of 1.0. Short CID sequence tags determined at a 3-ppm accuracy (**Figures S2B** and **S3B**) can nevertheless contribute valuable information by confirming ECD de novo assignments.

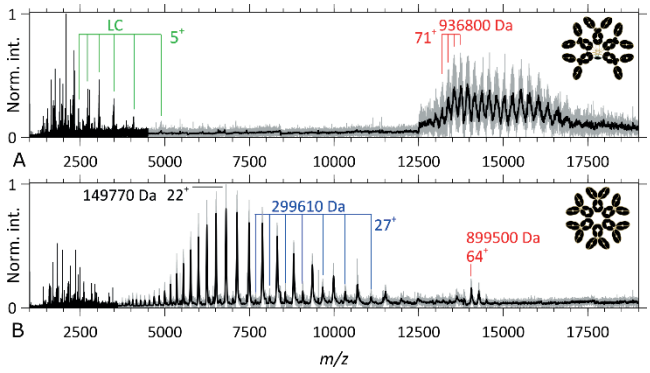


Figure 2 | Native top-down CID MS2 spectra of IgM- and IgG-based oligomeric immunoglobulins. (A) CID MS2 spectrum of the aWTA IgM precursor ion signal (z ~ 74) overlapping with high m/z product ions in the upper m/z range, while isotopically resolved intact LCs and backbone fragments are detected below m/z 5000. **(B)** CID MS2 spectrum of the aCD52 hexameric IgG1 precursor ion (z ~ 66) at high m/z, intact IgG1 monomers, and lower abundant IgG dimers in the 3500 ≤ m/z ≤ 12,500 range, and backbone fragments below m/z 3500. A Gaussian filter has been applied to the high m/z range with the resulting spectrum overlapping the raw data (in gray).

Native Top-Down MS of Oligomeric Immunoglobulins by ECD

While CID of large protein assemblies mostly results in subunit ejection, ECD is known to preferentially generate backbone fragments[60]. Practically, the addition of an electron to a cationic protein increases its internal energy by 6-7 eV of which about 3 eV is used to cleave the backbone and form (c, $z\cdot$)-ion pairs, while the residual energy contributes to fragment separation (the intermolecular hydrogen bond energies range from -6 to -3 kJ/mol or -0.06 to -0.03 eV)[61,62]. The efficiency of fragment generation by ECD can however be low, especially for large protein assemblies, as precursor ions are known to capture up to tens of electrons without fragmenting, a process sometimes referred to as ECnoD[23]. Applied to the precursor ions of the aWTA IgM and aCD52 hexameric IgG1, ECD yields precursor ions having captured tens of electrons, as shown in **Figures 3A, B** and **S4**, as well as **Figures 3C** and **S10** (m/z regions above the original ion signals). These ECnoD ions are unfortunately not very useful for sequencing and structural elucidation. Gratifyingly, in the 500 < m/z < 3000 region, a variety of fragment ions are observed, which are the result of ECD cleavages in the backbone of the immunoglobulin oligomers, as discussed below.

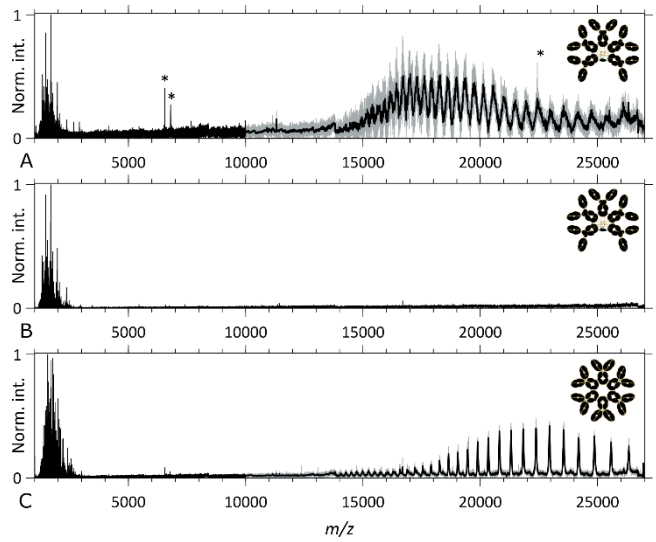


Figure 3 | Native top-down ECD MS2 spectra of IgM- and IgG-based oligomeric immunoglobulins. (A, B) ECD MS2 spectra of the aWTA IgM assembly and (C) ECD MS2 spectra of the aCD52 hexameric IgG1 assembly. **(A, C)** ECD MS2 spectra taken with the "Analyzer CE-Inject (V) UHMR" at its default value of 3200, displaying the precursor ion signals and the ECnoD product ions at high *m/z* and backbone fragments below *m/z* 3000, respectively, for the aWTA IgM and the aCD52 hexameric IgG1 assemblies. **(B)** ECD MS2 spectrum of the aWTA IgM assembly taken with the "Analyzer CE-Inject (V) UHMR" at a value of 3700, optimized for "low *m/z*" fragment ion detection. * denotes noise peaks.

Optimizing ECD on a Q Exactive UHMR requires reducing the kinetic energy of the ions in order to maximize residence time in the ECD cell to facilitate efficient electron capture. A significant fraction of the kinetic energy is transferred to the ions by the electrospray process and the drag force exerted by the gas entering the instrument can be removed by

performing extensive in-source trapping. In our setup, in-source trapping also serves to increase the internal energy of the precursor ions. We determined that by using a desolvation potential of -100 V, ECD fragment abundances were maximized for compounds ranging from Fab, F(ab')$_2$, and intact Igs to the here-studied close-to-MDa oligomeric Igs. Likely two reasons are contributing to this observation. First, as described by Loo and co-workers, ion-pair/salt-bridge rearrangements can occur upon collisional activation[59], which in turn impacts the hydrogen bonding interactions that contribute some of the radical hydrogens ultimately responsible for backbone cleavages and the more efficient formation of *c* and *z*· fragments[63]. Second, by increasing the internal energy of the ions, we bring them closer to the dissociation threshold of the noncovalent interactions stabilizing the *c* and *z*· ion pairs and thereby facilitating their separation[21,44].

Upon release from the source trap, minimized potential differences transfer the precursor ions first to the quadrupole – where they are mass-selected – then to the ECD cell where they undergo electron capture. Following their transmission through the ECD cell and C-trap, the ions – at this stage, a mixture of precursors and fragments – are trapped in the HCD cell. Reduced kinetic energies and pressure allow efficient trapping of fragments with the m/z up to a few thousand, while the precursor ions and the high m/z product ions responsible for the third harmonic interference, interfering with the low m/z fragment signal, are suppressed. Orbitrap measurements of large protein assemblies in the hundreds-of-kDa range are known to yield spurious frequencies, of which the third harmonic is the most problematic one as it overlaps, for precursors with m/z ranging from 10,000 to 30,000, with fragments about the 1-3 kDa range (**Figure S13**). High m/z signal suppression at low pressure is a consequence of the limited stopping power of the low-pressure buffer gas used to trap ions in the HCD cell.

In **Figure 3A**, we demonstrate that high-quality ECD fragment signals can be obtained for an aWTA IgM while retaining the detection of the charge-reduced precursor. Upon decreasing the analyzer's "CE-Inject (V) UHMR" setting from 3700 to 3200, we obtain the spectra as shown in **Figure 3B** with significantly improved signal-to-noise ratio and complete suppression of the precursor signal and its third harmonics. Charge deconvolution of the spectra of **Figure 3A, B** yield the $m/[z = 1]$ spectra displayed in **Figures** (**S5**, **S6**) and (**4**, **S8**, and **S9**), respectively.

Remarkably, similar to other monomeric Ig types and their subclasses, ECD of these very large ~ 1 MDa oligomeric Igs also results in backbone cleavages primarily in the LC and HC sequence segments outside of the disulfide loops that bridge their variable and CL/CHμ1 regions. For both the HC and LC, the sequence tags cover close to the whole range outside the disulfide loops with the gaps mostly filled, in the LC case, by *a*-ions. This remarkable feature illustrates that even for a ~ 930 kDa "heterohexamer" consisting of five intact IgMs and one J-chain, all extensively and covalently linked with each other in the $(IgM)_5$:$(J)_1$ assembly, native top-down ECD can produce informative sequence tags from the variable regions of IgMs. The $(IgM)_5$:$(J)_1$ assembly studied here was recombinantly produced by coexpressing the IgM and J chains together, whereafter the full assembly was purified. By expressing IgM without coexpression of the J-chain (data not shown), IgM is formed in a variety of oligomeric states, with the tetramer, pentamer,

and hexamer being the most abundant[57] and also with quite some IgM monomers expressed. Using this sample, we were able to produce and study, by native top-down ECD, the fragmentation behavior of the ~ 184 kDa aWTA IgM monomer, that is, having exactly the same sequence as the aWTA IgM present in the here-studied ~ 935 kDa J-chain-linked pentamer. Comparison of the native top-down ECD $m/[z = 1]$ spectra for the intact aWTA IgM with the J-chain and the aWTA IgM monomer (**Figure 4**, top and bottom spectra) reveals the high similarity between these ECD spectra. Notably, here, the precursor ions were quite different, that is, the 72-60 charge envelope of the ~ 935 kDa and the 30-25 charge envelope of the ~ 184 kDa IgM monomer. Correlation values between these spectra are over 70%. It is to be noted that the monomer was sprayed without making use of in-source trapping but relying on a "source fragmentation (eV)" setting of 120 instead.

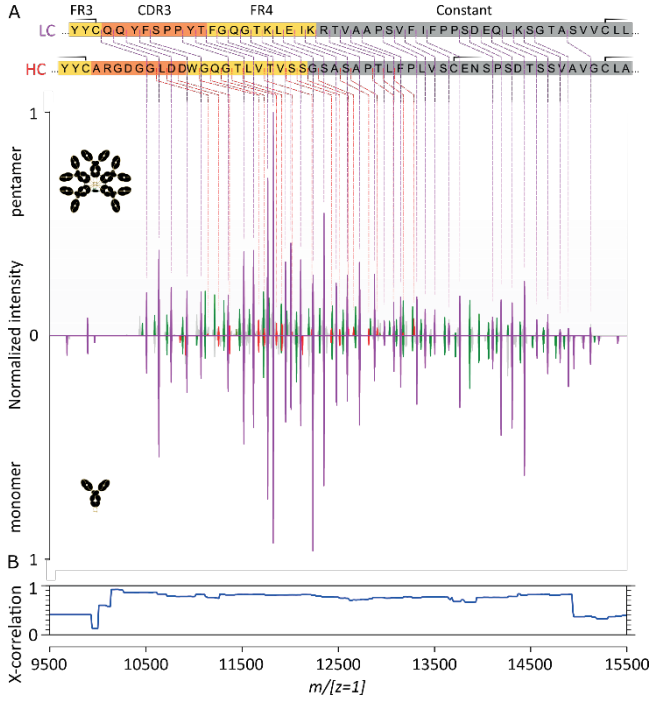


Figure 4 | Comparison of native top-down ECD fragment ion spectra of $(IgM)_5:(J)_1$ and the corresponding monomeric IgM. (A) Charge-deconvoluted ($m/[z = 1]$) native top-down ECD spectra of the aWTA IgM pentamer with the J-chain and the aWTA IgM monomer ("Analyzer CE-Inject (V) UHMR" at 3700 in both cases). Note that the sequence coverage of the HC stops before the cysteine bridging to the LC (the second of the three cysteine parts of the displayed HC sequence segment). HC *c*-ions are in red, LC *c*-ions are in purple, and the green peaks correspond to LC *a*-ions. **(B)** Cross-correlation analysis of the two native top-down ECD fragment ion spectra.

In **Figure 5**, the results of a similar approach are shown comparing the ECD spectra of the aCD52 IgG1-RGY hexamer and monomer. As in the case of the aWTA IgM with the J-chain, high-quality ECD fragment signals can be obtained for the aCD52 IgG1-RGY hexamer. Charge deconvolution of the ECD fragment spectra yields ECD $m/[z = 1]$ spectra characterized by nearly complete sequence coverage of both the LC and HC sequence

segments outside the disulfide loops bridging the variable and the constant regions (**Figures 5**, **S11**, and **S12**). For both the HC and LC, *a*-ions fill most of the gaps in the sequence tag, as observed previously[45,46].

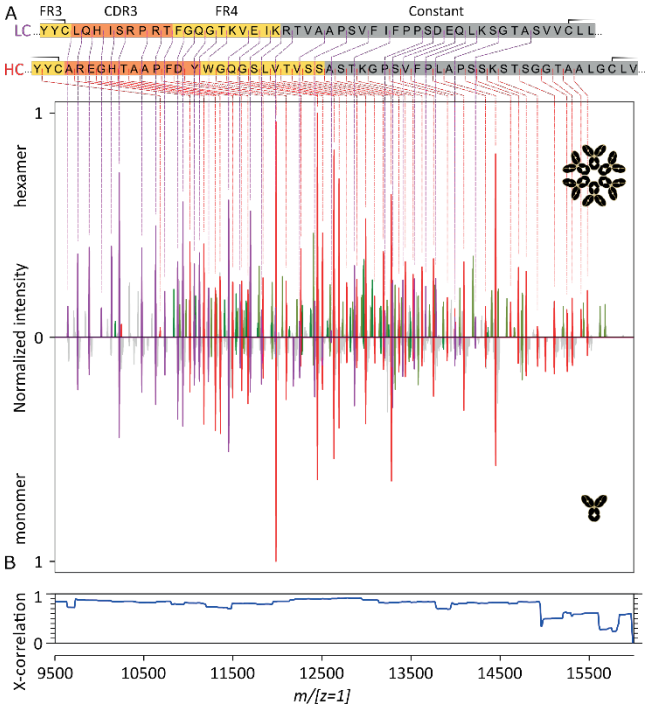


Figure 5 | Comparison of the native top-down ECD fragment ion spectra of $(IgG)_6$ and the corresponding monomeric IgG. (A) Charge-deconvoluted ($m/[z = 1]$) ECD spectra of the aCD52 IgG1 hexamer and its monomer ("Analyzer CE-Inject (V) UHMR" at 3200 in both cases). HC *c*-ions are in red, LC *c*-ions are in purple, and the green peaks correspond to LC *a*-ions. **(B)** Cross-correlation of the two native top-down ECD fragment ion spectra.

Comparison of the ECD $m/[z = 1]$ spectra for the intact aCD52 IgG1-RGY hexamer and a monomer (**Figure 5**, top and bottom spectra) establishes again the high similarity between these ECD spectra and the absence of an effect of the IgG oligomerization on the obtained sequence coverage. Here also, the precursor ions were quite different, that is, the 78-67 charge envelope of the ~ 890 kDa and the 29-23 charge envelope of the ~ 149 kDa IgG1-RGY monomer. Correlation values between the two spectra are over 80%. It is to be noted that the monomer was readily obtained from an aCD52 IgG1-RGY hexamer sample (**Figure 1**), and the acquisition was performed under identical conditions.

Conclusions

Over the last decades, proteomics has expanded from the traditional peptide-centric approaches (*i.e.*, bottom-up or shotgun proteomics with precursor masses typically being 500 < M_w < 5000 Da) to the direct analysis of intact proteins by the so-called top-down proteomics (with precursor masses typically in the range 5 kDa <M_w < 50 kDa). In the past years, this trend has been expanding even further to cover much larger protein assemblies by the native top-down proteomics (with precursor masses typically in the range 50 kDa < M_w < 0.5 MDa). These larger and more complex precursors require dedicated approaches to separate, fragment, and characterize analytes, tackled by native top-down proteomics. Here, we successfully demonstrate that native top-down proteomics can be extended to the important class of oligomeric immunoglobulins, with MWs reaching close to 1 MDa, and that ECD implemented on an Orbitrap UHMR can yield valuable sequence information to characterize and partially identify these important immune molecules. Specifically, the present approach allowed us to get straightforward-to-read isotopically resolved amino acid sequence ladders of *c*-ions (9000-16,000 Da) spanning the CDR3s and FR4s. The analysis of the intact oligomers was validated by performing ECD of their monomers. Only small differences (< 20%) were observed between the native top-down ECD spectra of the intact oligomers and their monomers, suggesting a low impact of the assembly on fragment formation.

Overall, we have demonstrated that the complexity and size of the pentameric J-chain-linked IgM and hexameric IgG1 structures do not adversely impact their ECD compared to smaller-monomeric Igs. As in the case of IgGs and IgAs[45,46], we showed that ECD without supplemental activation could be deliberately restricted to the highly informative hypervariable regions outside the disulfide loops of both the LC and HC. ECD without supplemental activation thereby proved to be uniquely suited to the screening of all isotypes and variants of immunoglobulins. This, in turn, may impact regulatory requirements for the characterization of therapeutic antibodies ranging from IgGs, IgAs, to IgMs, or, in a diagnostic context, facilitate the screening of complex mixtures in combination with orthogonal separation methods.

Notes

The authors declare the following competing financial interest(s): J.C. is an employee of MSVision, a company that offers instrument modifications as a service, and installs ECD modules in a variety of mass analyzers. All other authors have no conflict of interest to report.

Acknowledgements

We thank the members of the Heck laboratory for general support, especially Arjan Barendregt. We acknowledge the group of Janine Schuurman at Genmab for providing the recombinant RGY-mutant IgG samples and the Suzan Rooijakkers group at the Medical

Microbiology Department, University Medical Center Utrecht (The Netherlands), in particular, Piet Aerts and Carla Gosselaar-de Haas, for providing the recombinant IgM samples. This research received funding through the Netherlands Organization for Scientific Research (NWO) TTW project 15575 (Structural analysis and position-resolved imaging of macromolecular structures using novel mass spectrometry-based approaches), the ENPPS.LIFT.019.001 project (A.J.R.H., S.-H.L., and J.-F.G.), the NACTAR project 16442 (A.J.R.H. and M.A.d.B.), and the Spinoza Award SPI.2017.028 to A.J.R.H. This project received additional funding from the European Union's Horizon 2020 research and innovation program under grant agreement 686547 (EPIC-XS) for A.J.R.H.

References

1. Klingler, J. et al. Role of Immunoglobulin M and A Antibodies in the Neutralization of Severe Acute Respiratory Syndrome Coronavirus 2. *The Journal of Infectious Diseases* **223**, 957-970 (2020).
2. Michaud, E., Mastrandrea, C., Rochereau, N. & Paul, S. Human Secretory IgM: An Elusive Player in Mucosal Immunity. *Trends Immunol.* **41**, 141-156 (2020).
3. Ehrenstein, M.R. & Notley, C.A. The importance of natural IgM: scavenger, protector and regulator. *Nature Reviews Immunology* **10**, 778-786 (2010).
4. Brändlein, S. et al. Natural IgM Antibodies and Immunosurveillance Mechanisms against Epithelial Cancer Cells in Humans. *Cancer Res.* **63**, 7995-8005 (2003).
5. Vollmers, H.P. & Brändlein, S. Natural IgM antibodies: from parias to parvenus. *Histol. Histopathol.* **21**, 1355-1366 (2006).
6. Kretschmer, A., Schwanbeck, R., Valerius, T. & Rösner, T. Antibody Isotypes for Tumor Immunotherapy. *Transfusion Medicine and Hemotherapy* **44**, 320-326 (2017).
7. Hiramoto, E. et al. The IgM pentamer is an asymmetric pentagon with an open groove that binds the AIM protein. *Science Advances* **4**, eaau1199 (2018).
8. Kumar, N., Arthur, C.P., Ciferri, C. & Matsumoto, M.L. Structure of the human secretory immunoglobulin M core. *Structure* **29**, 564-571.e563 (2021).
9. Li, Y. et al. Structural insights into immunoglobulin M. *Science* **367**, 1014-1017 (2020).
10. Vollmers, H.P. & Brändlein, S. Natural IgM antibodies: The orphaned molecules in immune surveillance. *Adv. Drug Del. Rev.* **58**, 755-765 (2006).
11. Perkins, S.J., Nealis, A.S., Sutton, B.J. & Feinstein, A. Solution structure of human and mouse immunoglobulin M by synchrotron X-ray scattering and molecular graphics modelling: A possible mechanism for complement activation. *J. Mol. Biol.* **221**, 1345-1366 (1991).
12. Reddy, P.S. & Corley, R.B. The contribution of ER quality control to the biologic functions of secretory IgM. *Immunol. Today* **20**, 582-588 (1999).
13. Johansen, Braathen & Brandtzaeg Role of J Chain in Secretory Immunoglobulin Formation. *Scand. J. Immunol.* **52**, 240-248 (2000).
14. de Jong, R.N. et al. A Novel Platform for the Potentiation of Therapeutic Antibodies Based on Antigen-Dependent Formation of IgG Hexamers at the Cell Surface. *PLoS Biol.* **14**, e1002344 (2016).
15. Smith, R.I., Coloma, M.J. & Morrison, S.L. Addition of a mu-tailpiece to IgG results in polymeric antibodies with enhanced effector functions including complement-mediated cytolysis by IgG4. *J. Immunol.* **154**, 2226-2236 (1995).
16. Diebolder, C.A. et al. Complement Is Activated by IgG Hexamers Assembled at the Cell Surface. *Science* **343**, 1260-1263 (2014).
17. Wang, G. et al. Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Mol. Cell* **63**, 135-145 (2016).
18. Lubbers, R. et al. Carbamylation reduces the capacity of IgG for hexamerization and complement activation. *Clinical and Experimental Immunology* **200**, 1-11 (2020).
19. Wei, B. et al. Fc galactosylation follows consecutive reaction kinetics and enhances immunoglobulin G hexamerization for complement activation. *MAbs* **13**, 1893427 (2021).
20. Cruz, A.R. et al. Staphylococcal protein A inhibits complement activation by interfering with IgG hexamer formation. *Proc. Natl. Acad. Sci.* **118**, e2016772118 (2021).

21. Zhang, H., Cui, W. & Gross, M.L. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett.* **588**, 308-317 (2014).
22. Li, H., Wolff, J.J., Van Orden, S.L. & Loo, J.A. Native Top-Down Electrospray Ionization-Mass Spectrometry of 158 kDa Protein Complex by High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **86**, 317-320 (2014).
23. Zhou, M. et al. Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chemical Science* **11**, 12918-12936 (2020).
24. Ives, A.N. et al. Using 10,000 Fragment Ions to Inform Scoring in Native Top-down Proteomics. *J. Am. Soc. Mass. Spectrom.* **31**, 1398-1409 (2020).
25. Konijnenberg, A. et al. Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment. *Protein Sci.* **24**, 1292-1300 (2015).
26. Hale, O.J. & Cooper, H.J. Native Mass Spectrometry Imaging and In Situ Top-Down Identification of Intact Proteins Directly from Tissue. *J. Am. Soc. Mass. Spectrom.* **31**, 2531-2537 (2020).
27. Skinner, O.S. et al. Top-down characterization of endogenous protein complexes with native proteomics. *Nat. Chem. Biol.* **14**, 36-41 (2018).
28. Snyder, D.T., Panczyk, E.M., Somogyi, A., Kaplan, D.A. & Wysocki, V. Simple and Minimally Invasive SID Devices for Native Mass Spectrometry. *Anal. Chem.* **92**, 11195-11203 (2020).
29. O'Brien, J.P., Li, W., Zhang, Y. & Brodbelt, J.S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Chemical Society* **136**, 12920-12928 (2014).
30. Bashyal, A., Sanders, J.D., Holden, D.D. & Brodbelt, J.S. Top-Down Analysis of Proteins in Low Charge States. *J. Am. Soc. Mass. Spectrom.* **30**, 704-717 (2019).
31. Greisch, J.-F. et al. Expanding the mass range for UVPD-based native top-down mass spectrometry. *Chemical Science* **10**, 7163-7171 (2019).
32. Greisch, J.-F., van der Laarse, S.A.M. & Heck, A.J.R. Enhancing Top-Down Analysis Using Chromophore-Assisted Infrared Multiphoton Dissociation from (Phospho)peptides to Protein Assemblies. *Anal. Chem.* **92**, 15506-15516 (2020).
33. Mehaffey, M.R., Xia, Q. & Brodbelt, J.S. Uniting Native Capillary Electrophoresis and Multistage Ultraviolet Photodissociation Mass Spectrometry for Online Separation and Characterization of Escherichia coli Ribosomal Proteins and Protein Complexes. *Anal. Chem.* **92**, 15202-15211 (2020).
34. Zhou, M., Liu, W. & Shaw, J.B. Charge Movement and Structural Changes in the Gas-Phase Unfolding of Multimeric Protein Complexes Captured by Native Top-Down Mass Spectrometry. *Anal. Chem.* **92**, 1788-1795 (2020).
35. Li, H., Nguyen, H.H., Ogorzalek Loo, R.R., Campuzano, I.D.G. & Loo, J.A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **10**, 139-148 (2018).
36. Srzentić, K. et al. Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **31**, 1783-1802 (2020).
37. Riley, N.M. & Coon, J.J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **90**, 40-64 (2018).
38. Breuker, K. & McLafferty, F.W. Native Electron Capture Dissociation for the Structural Characterization of Noncovalent Interactions in Native Cytochrome c. *Angew. Chem. Int. Ed.* **42**, 4900-4904 (2003).
39. Zubarev, R.A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **15**, 12-16 (2004).
40. Zhang, H., Cui, W., Wen, J., Blankenship, R.E. & Gross, M.L. Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly. *J. Am. Soc. Mass. Spectrom.* **21**, 1966-1968 (2010).
41. Gadkari, V.V. et al. Enhanced Collision Induced Unfolding and Electron Capture Dissociation of Native-like Protein Ions. *Anal. Chem.* **92**, 15489-15496 (2020).
42. Williams, J.P. et al. Top-Down Characterization of Denatured Proteins and Native Protein Complexes Using Electron Capture Dissociation Implemented within a Modified Ion Mobility-Mass Spectrometer. *Anal. Chem.* **92**, 3674-3681 (2020).
43. Zhang, J., Loo, R.R.O. & Loo, J.A. Structural Characterization of a Thrombin-Aptamer Complex by High Resolution Native Top-Down Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **28**, 1815-1822 (2017).

44. Lermyte, F., Valkenborg, D., Loo, J.A. & Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* **37**, 750-771 (2018).
45. den Boer, M.A., Greisch, J.-F., Tamara, S., Bondt, A. & Heck, A.J.R. Selectivity over coverage in de novo sequencing of IgGs. *Chemical Science* **11**, 11886-11896 (2020).
46. Greisch, J.-F. et al. Generating Informative Sequence Tags from Antigen-Binding Regions of Heavily Glycosylated IgA1 Antibodies by Native Top-Down Electron Capture Dissociation. *J. Am. Soc. Mass. Spectrom.* **32**, 1326-1335 (2021).
47. Mao, Y., Valeja, S.G., Rouse, J.C., Hendrickson, C.L. & Marshall, A.G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **85**, 4239-4246 (2013).
48. Shaw, J.B. et al. Sequencing Grade Tandem Mass Spectrometry for Top–Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Anal. Chem.* **90**, 10819-10827 (2018).
49. Shaw, J.B. et al. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **92**, 766-773 (2020).
50. Fort, K.L. et al. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **17**, 926-933 (2018).
51. Pacholarz, K.J. et al. Molecular Insights into the Thermal Stability of mAbs with Variable-Temperature Ion-Mobility Mass Spectrometry. *ChemBioChem* **17**, 46-51 (2016).
52. Chambers, M.C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918-920 (2012).
53. Klein, J., Carvalho, L. & Zaia, J. Application of network smoothing to glycan LC-MS profiling. *Bioinformatics* **34**, 3511-3518 (2018).
54. Lukauskas, J.K.h.S. (Zenodo, 2019).
55. Gibbons, C.W.B. (Computational Mass Spectrometry @ Pacific Northwest National Laboratory, 2020).
56. Park, J. et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **14**, 909-914 (2017).
57. Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **17**, 395-398 (2020).
58. Metwally, H., Duez, Q. & Konermann, L. Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry. *Anal. Chem.* **90**, 10069-10077 (2018).
59. Loo, R.R.O. & Loo, J.A. Salt Bridge Rearrangement (SaBRe) Explains the Dissociation Behavior of Noncovalent Complexes. *J. Am. Soc. Mass. Spectrom.* **27**, 975-990 (2016).
60. Zhurov, K.O., Fornelli, L., Wodrich, M.D., Laskay, Ü.A. & Tsybin, Y.O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **42**, 5014-5030 (2013).
61. Breuker, K., Oh, H., Lin, C., Carpenter, B.K. & McLafferty, F.W. Nonergodic and conformational control of the electron capture dissociation of protein cations. *Proc. Natl. Acad. Sci.* **101**, 14011-14016 (2004).
62. Wendler, K., Thar, J., Zahn, S. & Kirchner, B. Estimating the Hydrogen Bond Energy. *The Journal of Physical Chemistry A* **114**, 9529-9536 (2010).
63. Ganisl, B. & Breuker, K. Does Electron Capture Dissociation Cleave Protein Disulfide Bonds? *ChemistryOpen* **1**, 260-268 (2012).

Supporting Information

Supplementary Methods

Cross-correlation analysis

Spectral comparison of native top-down ECD spectra involved an in-house implementation of the cross-correlation approach pioneered for mass spectra by Yates and coworkers[1,2]. Cross-correlation provides a metric for the pairwise comparison of spectra. The approach is ideally suited for cases where fragment isotope envelopes expand well beyond the first isotopic peak, and this peak's intensity is diminished relative to the base peak in the envelope. For charge-deconvoluted spectra, we report the normalized cross-correlation values for a 1000 *m/z* mass window slid over the spectra in steps of 10 *m/z*.

Supplementary Figures and Tables

Table S1 | Typical instrument transmission and ECD cell parameters for IgM and hexameric IgG1 RGY

Instrument

S-Lens RF Level	200	**IST (V)**	-100	**Skimmer Volt. (V)**	15.0
Inj. Flat. Offset (V)	3.0	**Bent. Flat. DC (V)**	2.0	**Gate Lens Volt. (V)**	2.13

ECD cell (V)

L1	2.12	**L2**	-24.2	**LM3**	7.0
L4	6.7	**FB**	1.5	**LM5**	7.0
L6	-26.2	**L7**	1.5		

Isolation windows (*m/z*)

IgM monomer	[5500-7300]	**IgG1 monomer**	5964 w 50
IgM pentamer	14,100 w 2500	**IgG1 hexamer**	12,266 w 2000

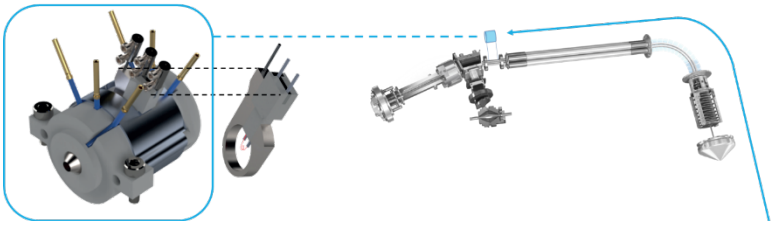


Figure S1 | Schematic of the ECD cell implementation.

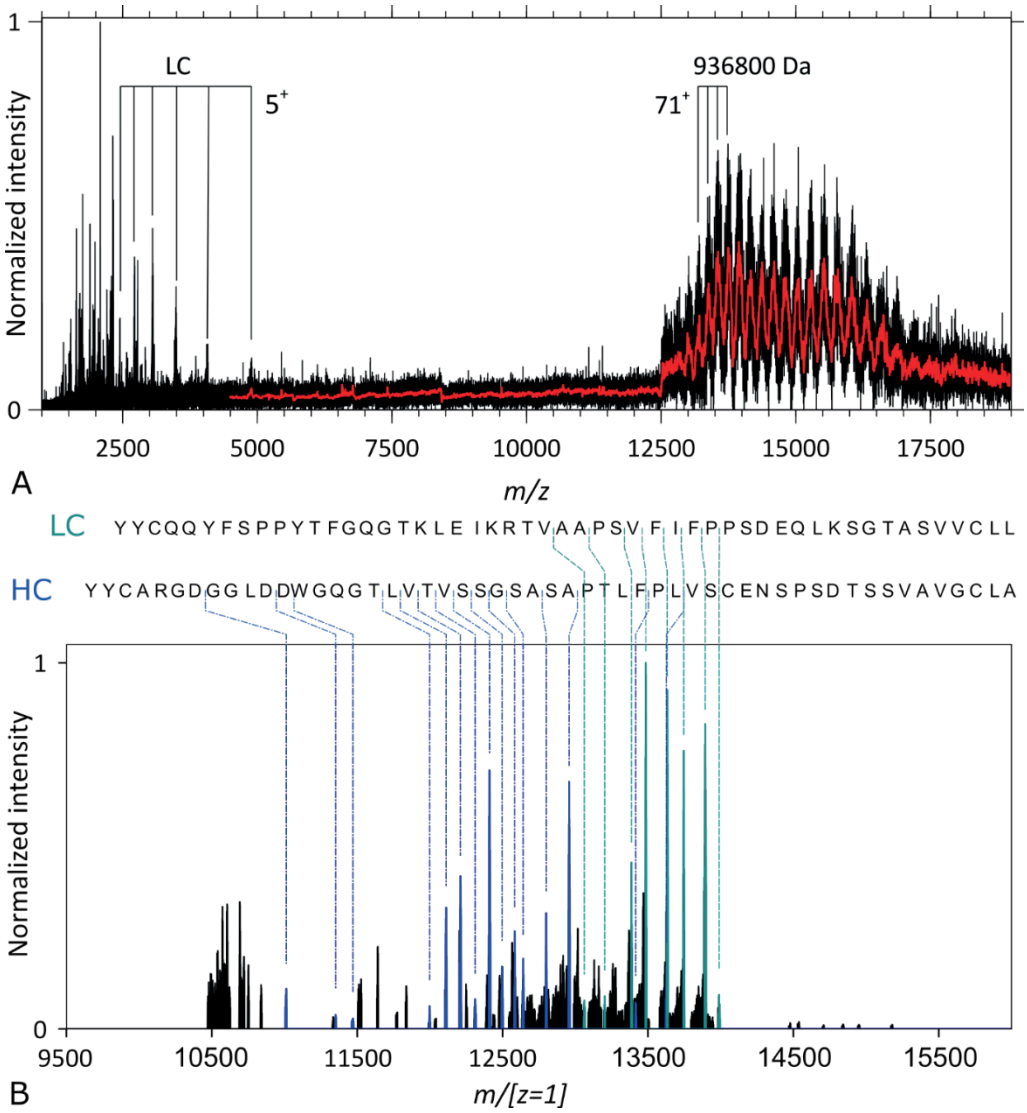


Figure S2 | Native top-down CID MS2 spectra of the aWTA IgM assembly. (A) Raw CID spectra with the precursor ion signal overlapping with high *m/z* product ions in the upper *m/z* range while in the region below *m/z* 5000 primarily isotopically resolved intact LCs and some backbone fragments are detected. The isolation window is centred on *m/z* 14,100 with a width of 2500. **(B)** Charge deconvoluted mass spectrum with LC and HC *b*-ion sequence tags highlighted in dark cyan and blue, respectively.

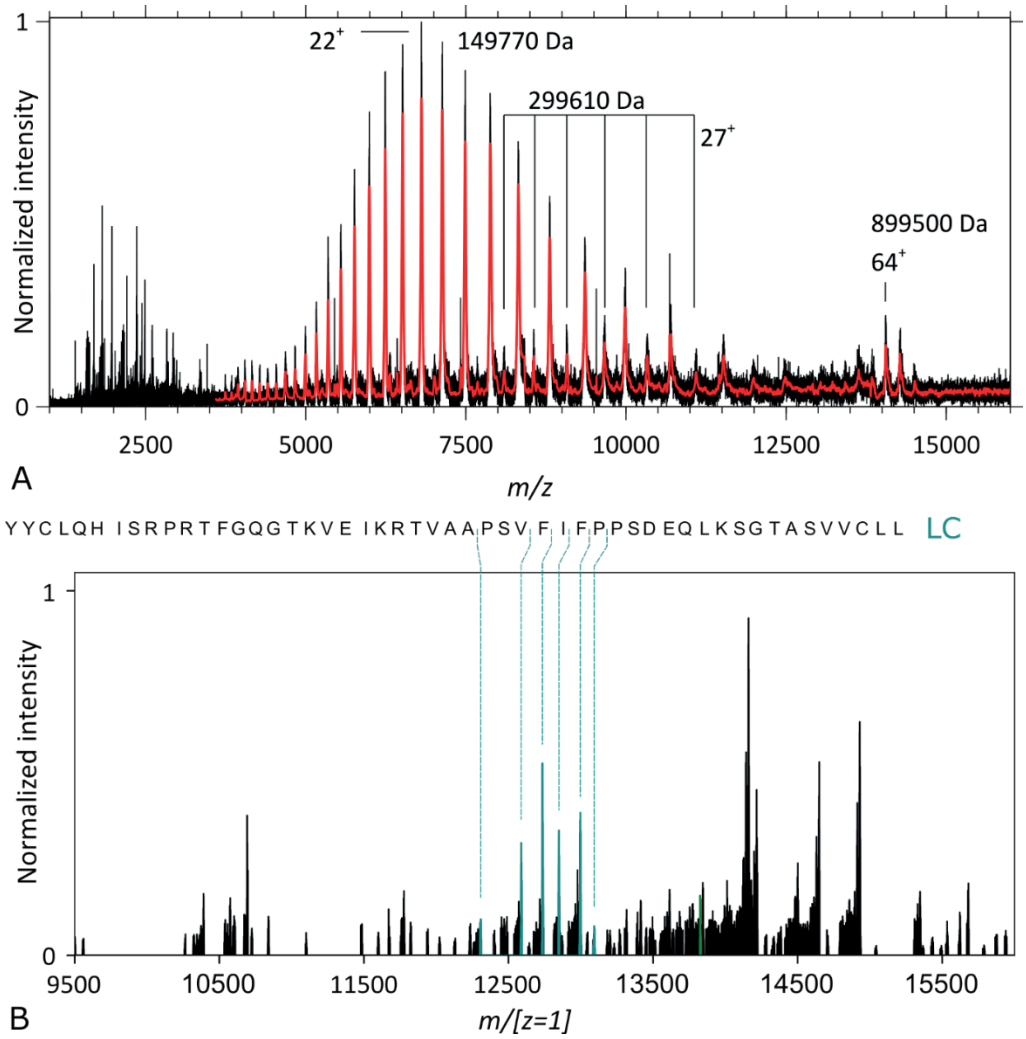


Figure S3 | Native top-down CID MS2 spectra of the aCD52 hexameric IgG1 assembly. (A) Raw CID spectra with the precursor ion signal at high *m/z*, intact IgG1 monomers in the 3500 ≤ *m/z* ≤ 12,500 range, and backbone fragments below *m/z* 3500. The isolation window extends from *m/z* 12,000 to 15,000. **(B)** Charge deconvoluted mass spectrum with a LC *b*-ion sequence tag highlighted in dark cyan.

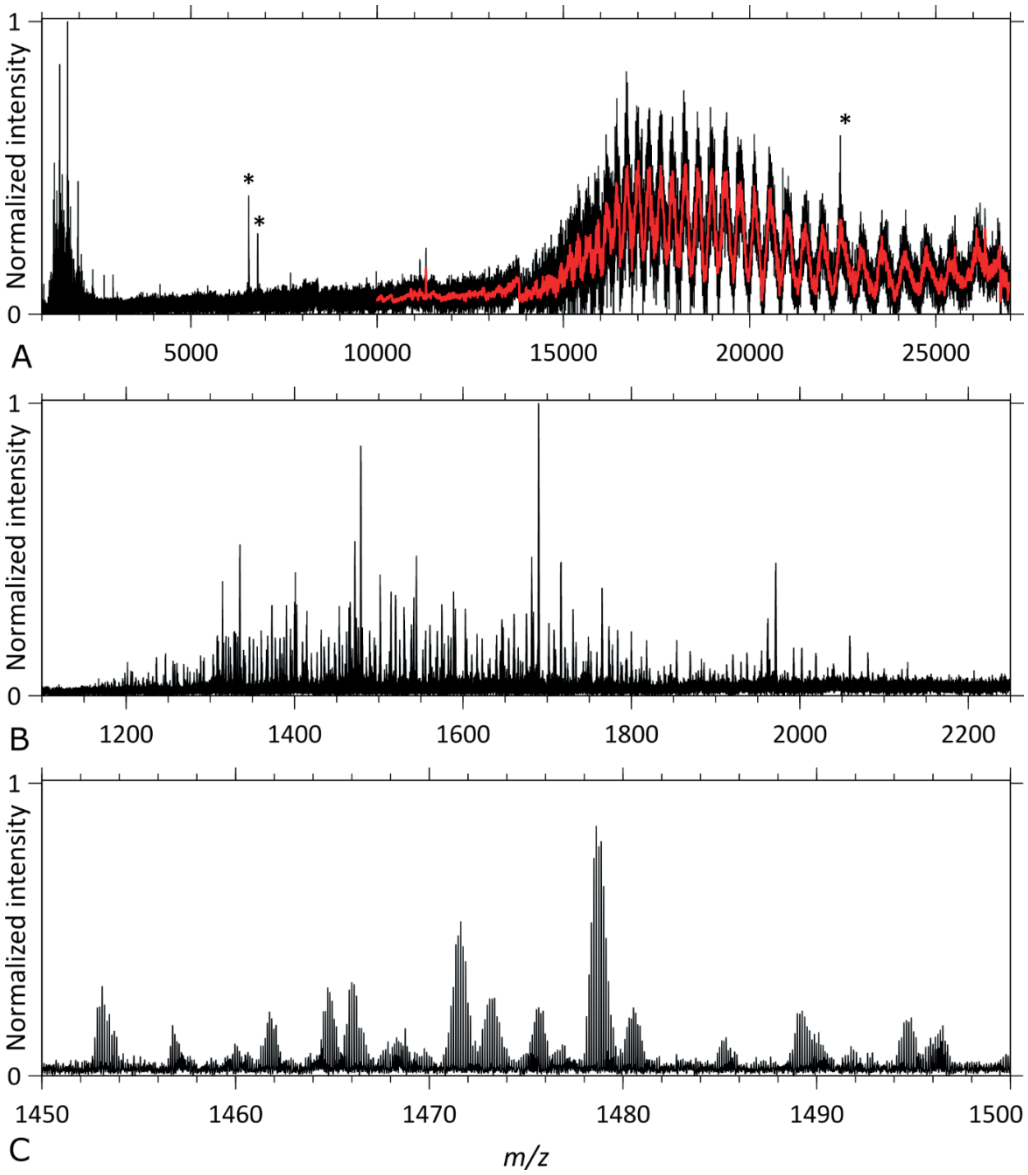


Figure S4 | Native top-down ECD MS2 spectra of the aWTA IgM assembly. All these ECD MS2 spectra were taken with the "Analyzer CE-Inject (V) UHMR" at its default value of 3200. **(A)** displaying the precursor ion signal at high *m/z*, and **(B)** backbone fragment ions below *m/z* 3000. C) Illustration of some of the isotopically resolved fragment ion signals.

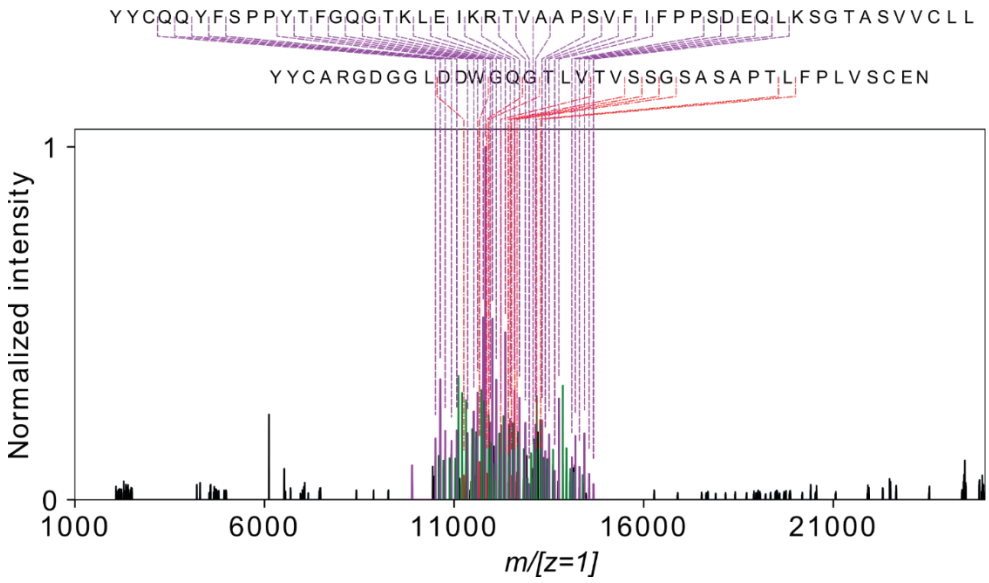


Figure S5 | Charge deconvoluted native top-down ECD MS2 spectra of the aWTA IgM assembly.

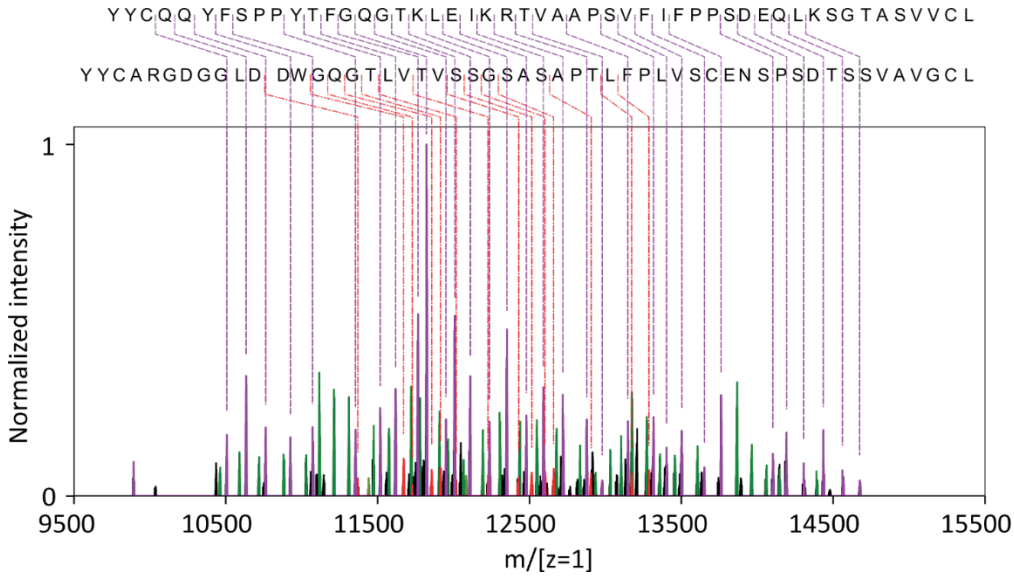


Figure S6 | Native top-down ECD MS2 spectra of the aWTA IgM assembly with the Analyzer CE-Inject (V) UHMR 3700.

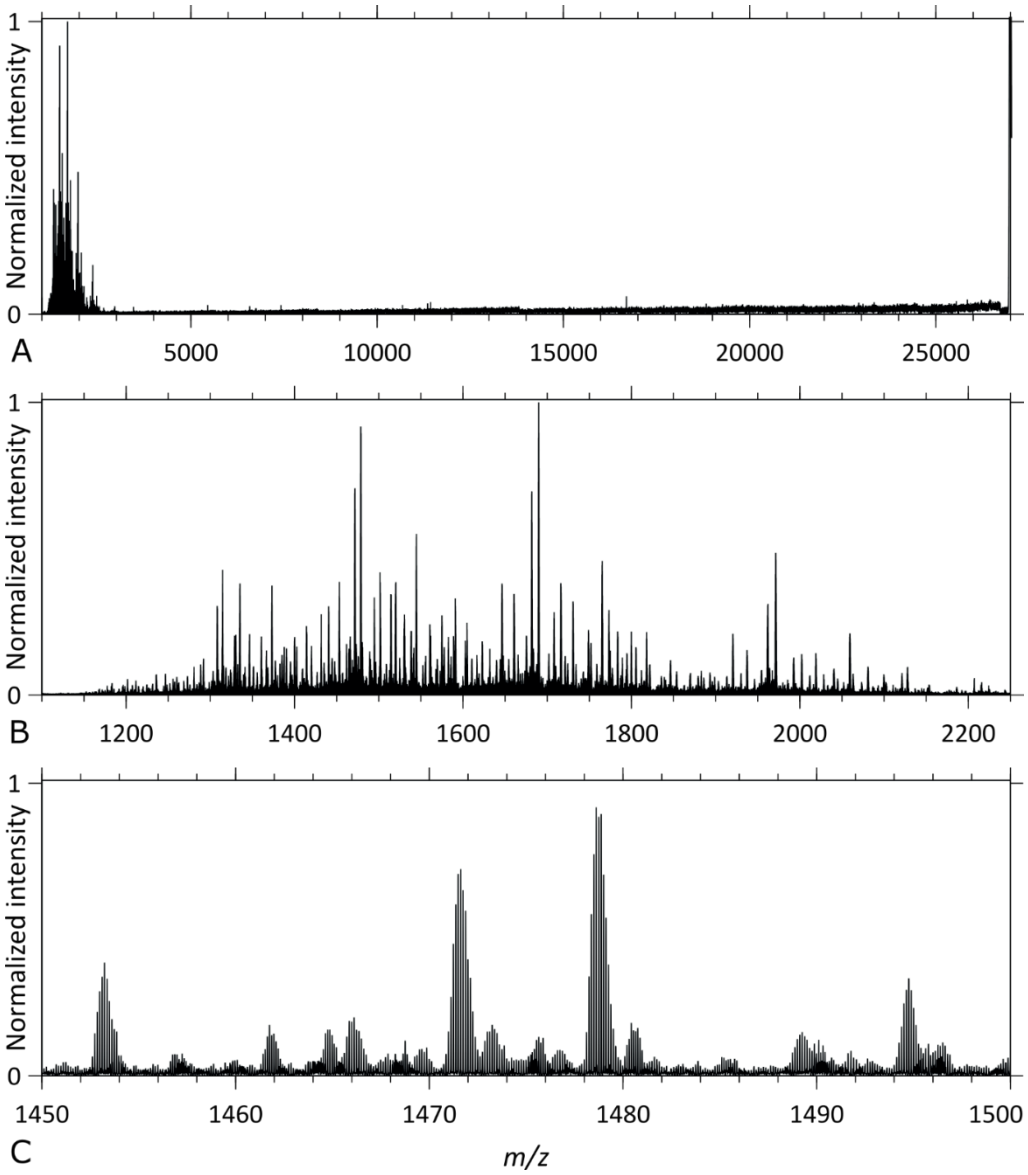


Figure S7 | Native top-down ECD MS2 spectra of the aWTA IgM assembly. All these ECD MS2 spectra of the aWTA IgM assembly were taken with the "Analyzer CE-Inject (V) UHMR" at a value of 3700, with **(A)** displaying the precursor ion signal at high *m/z*, and **(B)** backbone fragment ions below *m/z* 3000. **(C)** Illustration of some of the isotopically resolved fragment ion signals.

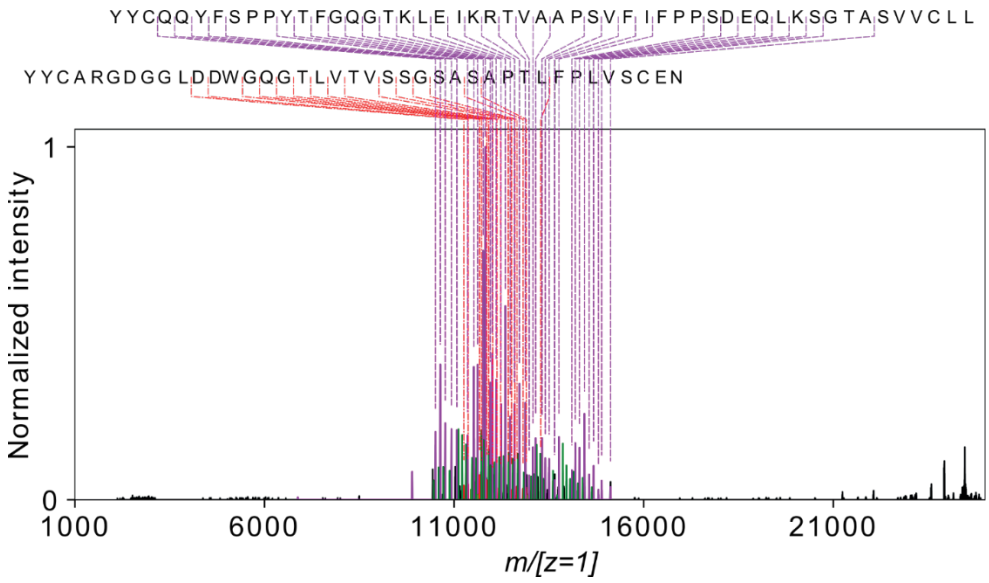


Figure S8 | Charge deconvoluted native top-down ECD MS2 spectra of the aWTA IgM assembly.

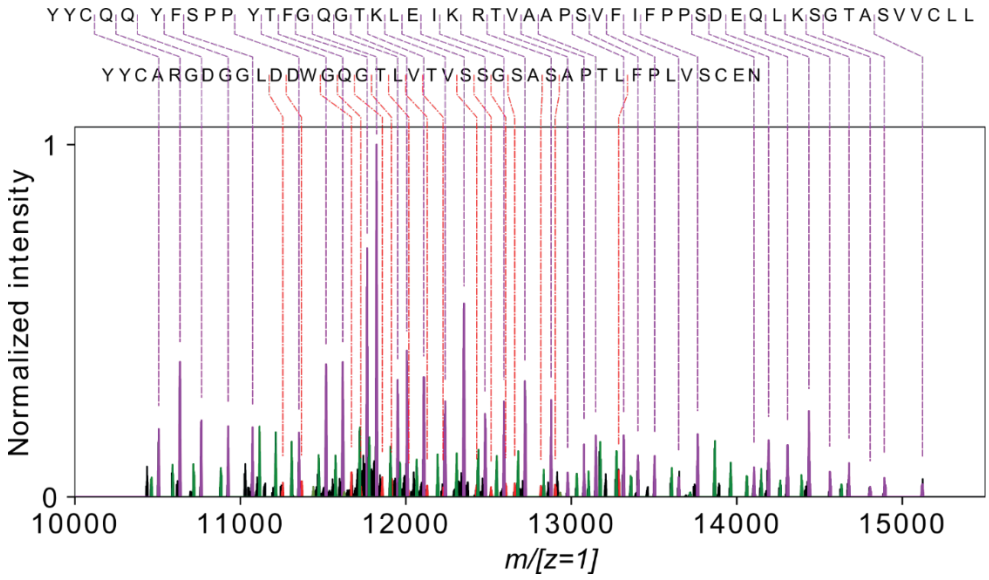


Figure S9 | Charge deconvoluted native top-down ECD MS2 spectra of the aWTA IgM assembly. Identical to corresponding figure part in main text.

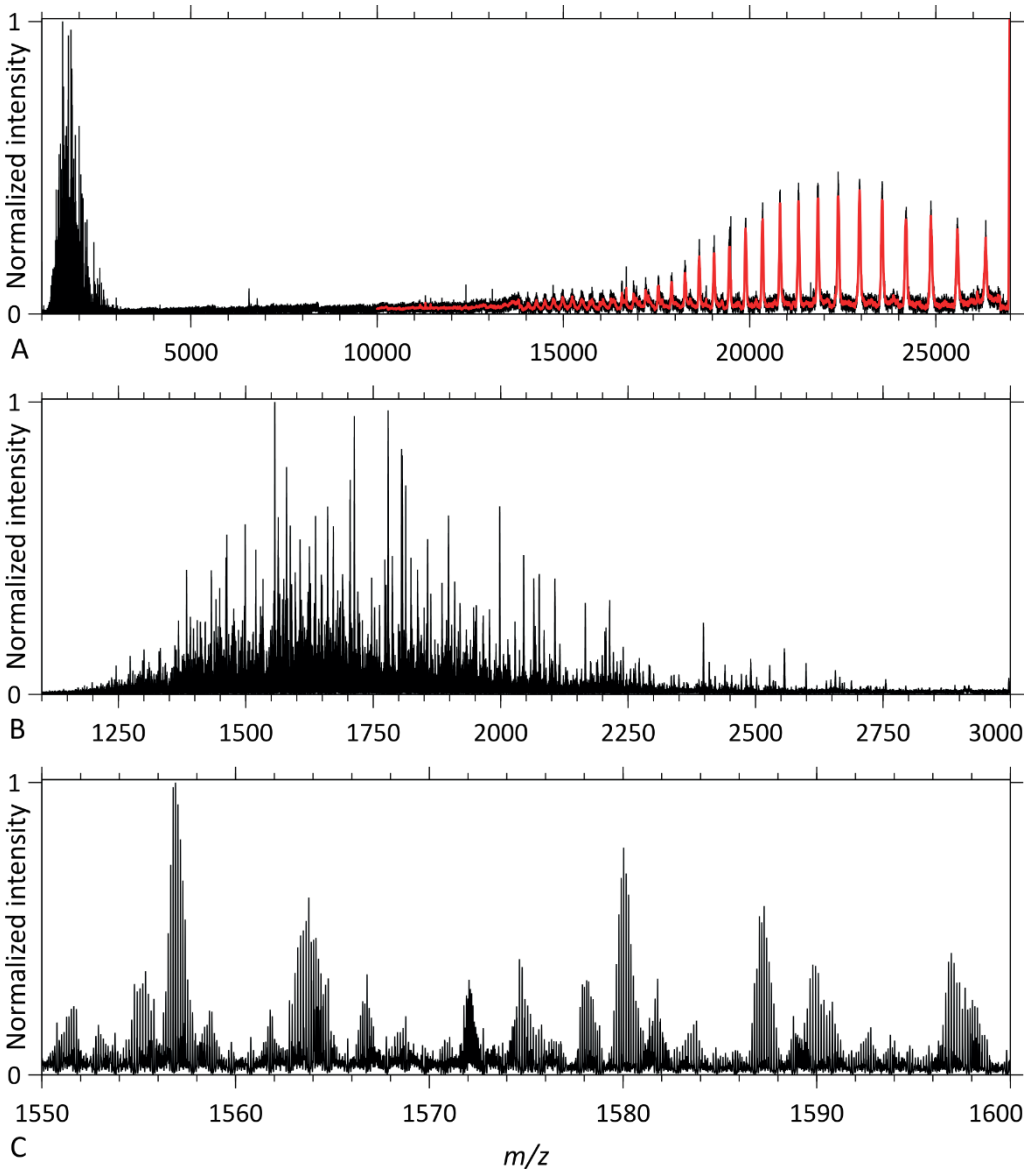


Figure S10 | Native top-down ECD MS2 spectra of the hexameric aCD52 IgG1 assembly. All these spectra were taken with the "Analyzer CE-Inject (V) UHMR" at its default value of 3200. **(A)** displaying the precursor ion signal at high *m/z*, and **(B)** backbone fragment ions below *m/z* 3000. **(C)** Illustration of some of the isotopically resolved fragment ion signals.

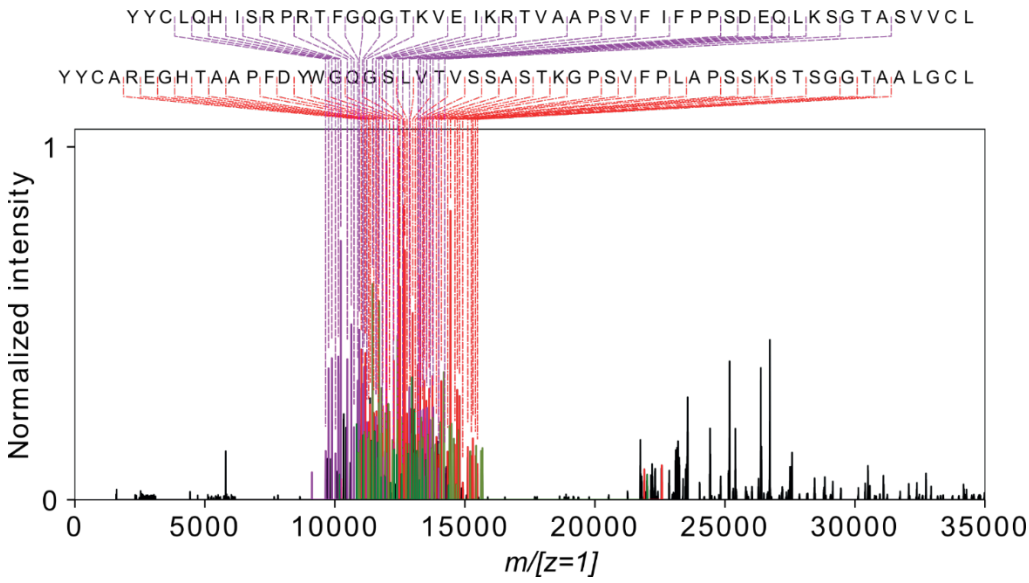


Figure S11 | Charge deconvoluted native top-down ECD MS2 spectra of the hexameric aCD52 IgG1 assembly.

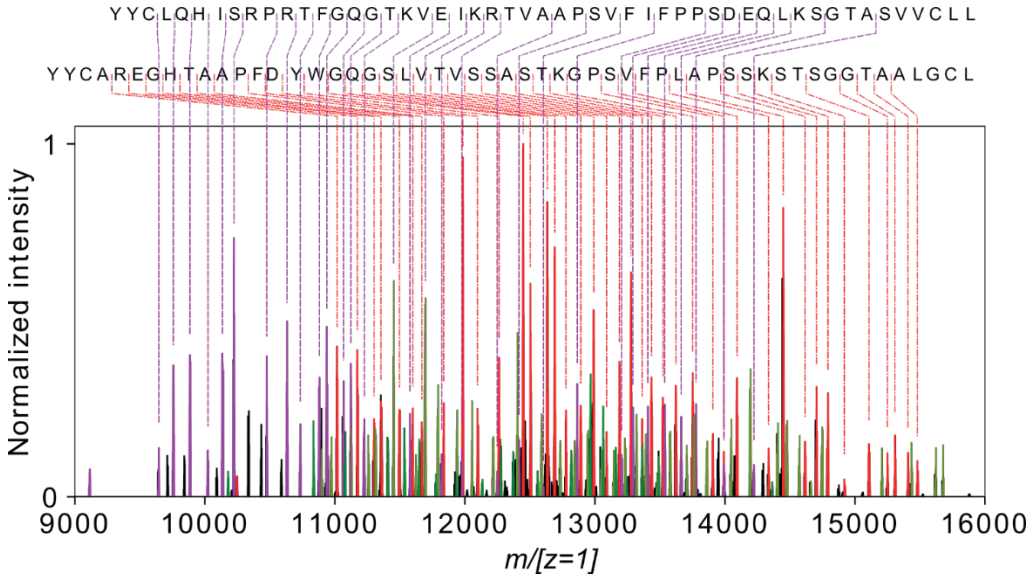


Figure S12 | Charge deconvoluted native top-down ECD MS2 spectra of the hexameric aCD52 IgG1 assembly. Identical to corresponding figure part in main text.

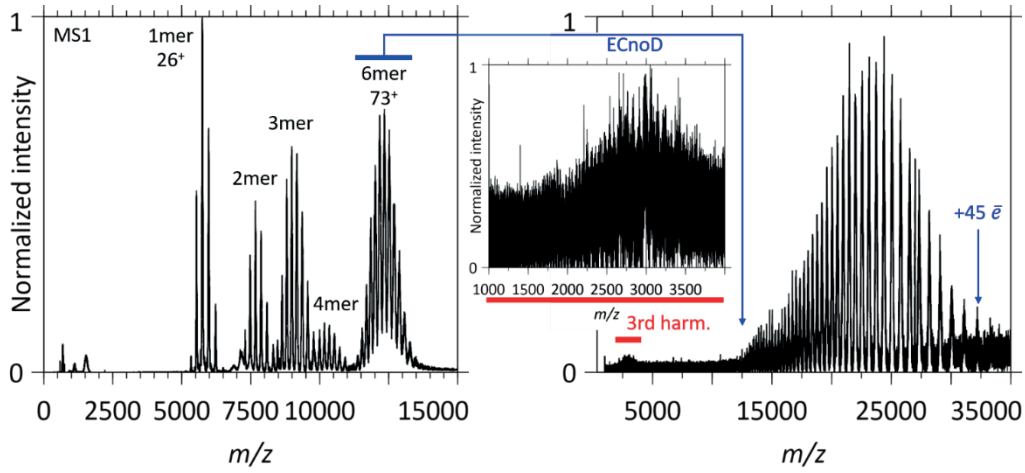


Figure S13 | MS1 of the hexameric aCD52 IgG1 solution (left). ECnoD spectrum of the hexameric aCD52 IgG1 assembly with the 3rd harmonic as inset (Right). Up to 45 electrons are captured by the hexamer.

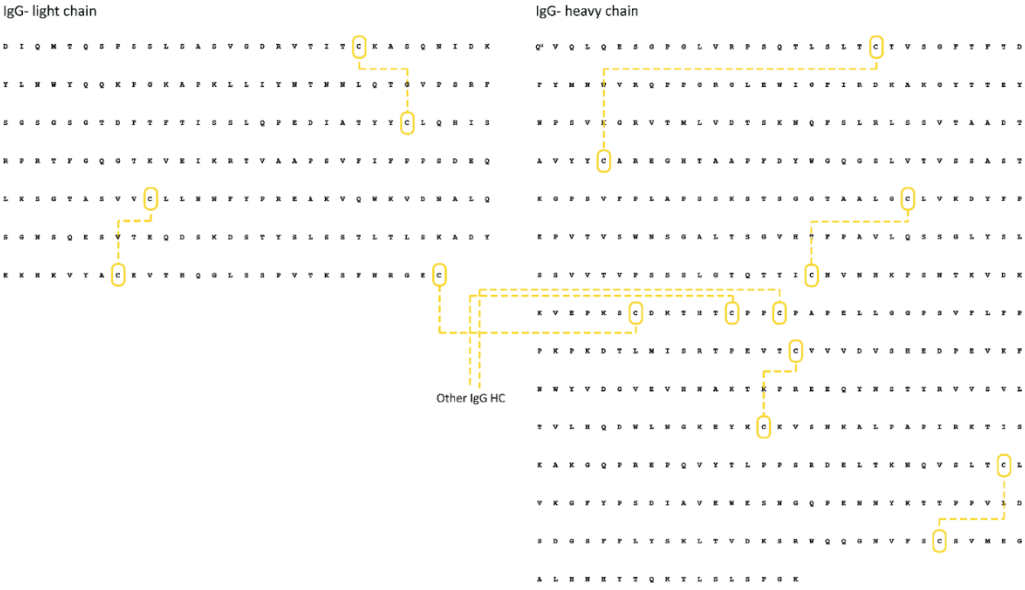


Figure S14 | Disulfide connectivity maps within IgG.

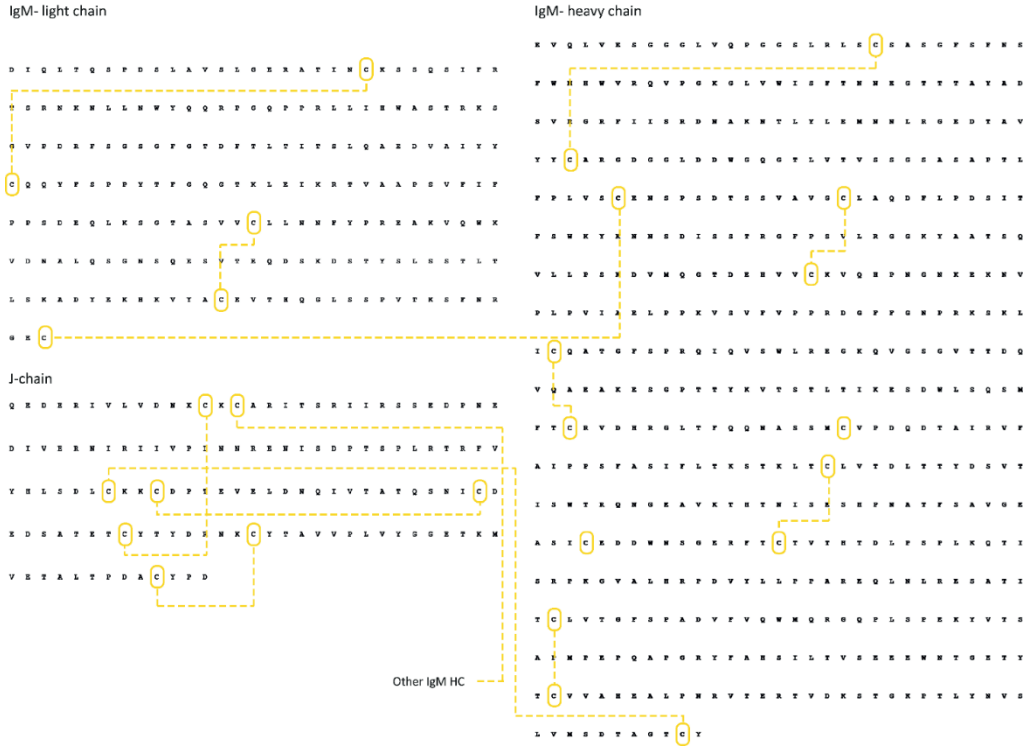


Figure S15 | Disulfide connectivity maps within IgM.

Data Access

All raw data files can be accessed at: https://doi.org/10.6084/m9.figshare.16873249.

Supplementary References

1. Venable, J.D., Xu, T., Cociorva, D. & Yates, J.R. Cross-Correlation Algorithm for Calculation of Peptide Molecular Weight from Tandem Mass Spectra. *Anal. Chem.* **78**, 1921-1929 (2006).
2. Greisch, J.-F. et al. Generating informative sequence tags from antigen-binding regions of heavily glycosylated IgA1 antibodies by native top-down electron capture dissociation. *J. Am. Soc. Mass. Spectrom.* **32**, 1326-1335 (2021).

CHAPTER 10

DIRECT MASS SPECTROMETRY-BASED DETECTION AND ANTIBODY SEQUENCING OF MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE FROM PATIENT SERUM: A CASE STUDY

Weiwei Peng[1,2], Maurits A. den Boer1[,2], Sem Tamara[1,2], Nadia Mokiem[1,2], Sjors P.A. van der Lans[3], Albert Bondt[1,2], Douwe Schulte[1,2], Pieter-Jan Haas[3], Monique C. Minnema[4], Suzan H.M. Rooijakkers[3], Arjan D. van Zuilen[5], Albert J.R. Heck[1,2], and Joost Snijder[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomic Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Medical Microbiology, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

[4] Department of Hematology, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

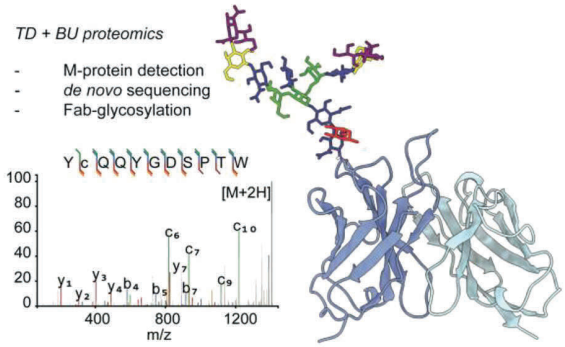
[5] Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

Published as:

Direct Mass Spectrometry-Based Detection and Antibody Sequencing of Monoclonal Gammopathy of Undetermined Significance from Patient Serum: A Case Study
Weiwei Peng, Maurits A. den Boer, Sem Tamara, Nadia J. Mokiem, Sjors P. A. van der Lans, Albert Bondt, Douwe Schulte, Pieter-Jan Haas, Monique C. Minnema, Suzan H. M. Rooijakkers, Arjan D. van Zuilen, Albert J. R. Heck, and Joost Snijder
J. Proteome Res. 2023, XXXX, XXX, XXX-XXX
DOI: doi.org/10.1021/acs.jproteome.3c00330

Abstract

Monoclonal gammopathy of undetermined significance (MGUS) is a plasma cell disorder, leading to the presence of monoclonal antibody (*i.e.*, M protein) in serum, without clinical symptoms. Here we present a case study in which we detect MGUS by liquid-chromatography coupled with mass spectrometry (LC-MS) profiling of IgG1 in human serum. We detected a Fab-glycosylated M protein and determined the full heavy and light chain sequences by bottom-up proteomics techniques using multiple proteases, further validated by top-down LC-MS. Moreover, the composition and location of the Fab glycan could be determined in CDR1 of the heavy chain. The outlined approach adds to an expanding mass spectrometry-based toolkit to characterize monoclonal gammopathies such as MGUS and multiple myeloma, with fine molecular detail. The ability to detect monoclonal gammopathies and determine M protein sequences straight from peripheral blood samples by mass spectrometry provides new opportunities to understand the molecular mechanisms of these diseases.



Introduction

Monoclonal gammopathy of undetermined significance (MGUS) is a plasma cell disorder, leading to the presence of a predominant monoclonal antibody (*i.e.*, an M protein) in patient serum[1]. MGUS is a preclinical stage of Multiple Myeloma (MM), with an estimated annual risk of 1% to progress to MM[2]. The most common antibody isotype in MGUS patients is IgG, which is a heterodimer consisting of two identical pairs of heavy chains (HC) and light chains (LC)[3,4]. All IgG share one conserved N-linked glycosylation site on each copy of the HC in the Fc region[5]. However, the M proteins present in both MGUS and MM patients have been reported to have a higher frequency of unusual additional glycosylation in the Fab region, present in the variable domains of either the light or heavy chains[6,7].

We have recently developed methods for direct mass spectrometry-based repertoire profiling and sequencing of IgG1 from human serum[8]. In this method, IgG is affinity purified from serum samples, followed by selective digestion and release of the IgG1 Fab portion[9]. Subsequent analysis of the released Fabs by reversed-phase liquid chromatography, coupled with mass spectrometry (LC-MS), establishes a highly resolved map of antibody clones separated by mass and retention time, spanning at least 2 orders of magnitude in abundance. By spiking in monoclonal antibodies at known concentrations, absolute concentrations of endogenous clones can also be estimated by normalizing their signal intensities. Typically, we detect a few hundred of the most abundant clones, together making up about 50-90% of the total subclass concentration. The most abundant IgG1 clones are generally in the order of 0.01-0.05 mg/mL, with sometimes outliers of up to about 1.0 mg/mL in hospitalized patients in critical condition[10,11].

While screening serum IgG1 repertoires of a new cohort of donors by LC-MS, we observed a donor whose repertoire was exceptionally dominated by a single IgG1 clone exhibiting a very high concentration of approximately 10 mg/mL in serum. This single clone contributes approximately 98% to the total amount of all IgG1 molecules in the serum of this patient. Subsequent clinical tests confirmed that this patient was affected by MGUS. We also found that the M protein harbored abundant Fab glycosylation. Combining the above IgG1 profiling method with bottom-up proteomics-based sequencing, we were able to recover the full sequence of the antibody heavy and light chains. The Fab glycosylation could be traced back to a specific residue in CDR1 of the heavy chain and the attached *N*-glycan structures could be assigned based on tandem MS spectra of the corresponding glycopeptides. This case illustrates how integrated bottom-up and top-down proteomics can be used to detect MGUS and other monoclonal gammopathies, sequence the associated monoclonal antibody in the presence of Fab glycosylation against a background of serum IgG1, and how the composition of this Fab glycosylation can be determined and localized from the same starting material.

Materials and Methods

Cohort and trial information

From 2015 through 2019, patients who underwent kidney transplantation were asked to participate in a biobank to evaluate immunological developments after kidney transplantation. All participants provided written informed consent to collect clinical data and serum samples pre-transplantation and at month 1, 3, 6, and 12 post-transplantation. The study was approved by the local Biobank Research Ethics Committee (protocol 15-019). Serum samples of patients with a recorded bacterial infection after kidney transplantation were identified and analyzed to evaluate the immunological response to such an infection. In one of the patients, the samples showed a few extremely abundant clones.

Serum IgG1 clonal profiling by LC-MS

IgG purification and Fab production. IgG1 clonal profiling was performed based on a method previously described by Bondt et al[8]. Two internal reference mAbs (trastuzumab and alemtuzumab) were spiked into serum samples of 10 µL to a final concentration of 20 µg/mL (200 ng), after which IgG was captured using 10 µL CaptureSelect FcXL affinity matrix (20 µL slurry, Thermo Fisher Scientific) in a spin column. After binding for 60 min on a shaker at 750 rpm and room temperature, columns were washed in four sequential rounds by adding 150 µL PBS and removing the liquid by centrifugation for 1 min at 500 g. IgG1 Fab molecules were released through on-bead proteolytic digestion using 100 U IgdE (FabALACTICA; Genovis) in 50 µL 150 mM sodium phosphate pH 7.0 overnight on a shaker at 750 rpm and 37 °C. Liquid containing free IgG1 Fabs was captured through centrifugation for 1 min at 1000 g.

LC-MS IgG1 clonal profiling. Analysis of IgG1 Fabs was performed by reversed-phase LC-MS using a Vanquish Flex UHPLC system (Thermo Fisher Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific). Chromatographic separation was performed on a 1 x 150 mm MAbPac column (Thermo Fisher Scientific) at 80 °C and using a flow rate of 150 µL/min. Mobile phase A consisted of MilliQ water with 0.1% formic acid, and mobile phase B of acetonitrile with 0.1% formic acid. Samples were run starting with a 10%-25% B ramp with the spray voltage turned off for 2 minutes to wash away salts. This was followed by a 54 min linear gradient of 25%-40% B, a 95% B wash and re-equilibration at 10% B. The mass spectrometer was operated at low pressure setting in Intact Protein mode at a set resolution of 7,500 at 200 *m/z.* For every scan, 5 µscans were acquired with an *m/z* range of 500 - 4000 using an AGC target of 300% with a maximum injection time of 50 ms. The RF lens was set to 40% and a source fragmentation energy of 15 V was applied. Raw data were processed by sliding window deconvolution using the ReSpect algorithm in BioPharma Finder v3.2 (Thermo Fisher Scientific). Further analysis was performed using an in-house Python library described by Bondt *et al.* Components with masses between 45,000 and 53,000 Da, most intense charge states above *m/z* 1000, and a Score of over 40 were considered valid Fab identifications.

Bottom-up proteomics

In-gel digestion. Fab (3 µg/lane) was loaded on a 4%-12% Bis-Tris precast gel (Bio-rad) in non-reducing conditions and run at 120 V in 3-Morpholinopropane-1-sulfonic acid (MOPS) buffer (Bio-rad). Bands were visualized with Imperial Protein Stain (Thermo Fisher Scientific), and the size of the fragments was evaluated by running a protein standard ladder (Bio-rad). The Fab bands were cut and reduced by 10 mM TCEP at 37 °C, then alkylated in 40 mM IAA at RT in the dark, followed by alkylation in 40 mM IAA at RT in the dark. The Fab bands were digested by trypsin, chymotrypsin, thermolysin, and alpha lytic protease at 37 °C overnight in 50 mM ammonium bicarbonate buffer. The peptides were extracted with two steps of incubation at RT in 50% ACN with 0.01% TFA, and then 100% ACN respectively. The peptides were dried in a SpeedVac. To obtain the sequence of the glycosylated Fab, the N-linked glycan was removed by PNGaseF at 37 °C overnight, followed by in-gel digestion as described above.

LC-MS/MS analysis. The digested peptides were separated by online reversed-phase chromatography on a Dionex UltiMate 3000 (Thermo Fisher Scientific) (column packed with Poroshell 120 EC C18; dimensions 50 cm × 75 µm, 2.7 µm, Agilent Technologies) coupled to an Orbitrap Fusion or Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). Samples were eluted over a 90 min gradient from 0 to 35% acetonitrile at a flow rate of 0.3 µL/min. Peptides were analyzed with a resolution setting of 60,000 in MS1. MS1 scans were obtained with a standard automatic gain control (AGC) target, a maximum injection time of 50 ms, and a scan range of 350-2000. The precursors were selected with a 3 *m/z* window and fragmented by stepped high-energy collision dissociation (HCD) and electron-transfer higher-energy collision dissociation (EThcD). The stepped HCD fragmentation included steps of 25, 35, and 50% normalized collision energies (NCE). EThcD fragmentation was performed with calibrated charge-dependent electron-transfer dissociation (ETD) parameters and 27% NCE supplemental activation. For both fragmentation types, MS2 scans were acquired at a 30,000 resolution, a 4e5 AGC target, a 250 ms maximum injection time, and a scan range of 120-3500.

Peptide sequencing from MS/MS spectra. MS/MS spectra were used to determine *de novo* peptide sequences using PEAKS Studio X (version 10.6)[12,13]. We used a tolerance of 20 ppm and 0.02 Da for MS1 and 0.02 Da for MS2, respectively. Carboxymethylation was set as a fixed modification of cysteine and a variable modification of peptide N-termini and lysine. Oxidation of methionine and tryptophan, pyroglutamic acid modification of N-terminal glutamic acid, and glutamine were set as additional variable modifications. The CSV file containing all the *de novo* sequenced peptides was exported for further analysis.

Template-based assembly via Stitch. Stitch[14] (1.1.2) was used for the template-based assembly. The human antibody database from IMGT was used as a template. The cutoff score for the *de novo* sequenced peptide was set as 90/70 and the cutoff score for the template matching was set as 10. All the peptides supporting the sequences were examined manually. The ions for annotation of the CDR regions were exported and visualized by Interactive Peptide Spectral Annotator[15].

Glycoproteomics data analysis. Chymotryptic digested peptides were used to search for site-specific glycosylation via Byonic (v5.0.3)[16]. The *de novo* obtained sequences were

selected as the protein database. Four missed cleavages were permitted using *C*-terminal cleavage at WFMLY for chymotrypsin. Carboxymethylation of cysteine was set as fixed modification, oxidation of methionine/tryptophan as variable rare 1, Gln to pyro-Glu and Glu to pyro-Glu on the *N*-terminus of protein as rare 1, and *N*-glycan modifications were set as variable rare 1. The *N*-glycan 132 human database from Byonic was applied in the search. All reported glycopeptides in the Byonic result files were manually inspected for quality of fragment assignments.

Middle-down proteomics

The reduced Fab was freshly prepared by incubating with TCEP at 60°C for 30 min before injecting to MS. Around 1 µg sample was used for a single measurement. Reduced Fab was measured by LC-MS/MS. Samples were loaded on a Vanquish Flex UHPLC system (Thermo Fisher Scientific), equipped with a 1 mm x 150 mm MAbPac RP analytical column (Thermo Fisher Scientific), directly coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). The samples were eluted over 22 min at a 150 µL/min flow rate. Gradient elution was achieved by using two mobile phases A (0.1% HCOOH in Milli-Q) and B (0.1% HCOOH in CH_3CN) and ramping up B from 10 to 25% over one minute, from 25 to 40% over 14 min, and from 40 to 95% over one minute. MS data were collected with the instrument operating in Intact Protein and Low Pressure mode. Spray voltage was set at 3.5 kV, capillary temperature 350 °C, probe heater temperature 100 °C, sheath gas flow 15, auxiliary gas flow 5, and source-induced dissociation was set at 15 V. Separate Fab chains were analyzed with a resolution setting of 120,000. MS1 scans were acquired in the *m/z* range 500-3,000 with the 250% AGC target and a maximum injection time set to either 50 ms for the 7,500 resolution or 250 ms for the 120,000 resolution. In MS1, 2 µscans were recorded for the 7,500 resolution and 5 µscans for the 120,000 resolution per scan. Data-dependent mode was defined by the number of scans: a single scan for intact Fabs and two scans for separate Fab chains. MS/MS scans were acquired with a resolution of 120,000, a maximum injection time of 500 ms, a 1000% AGC target, and 5 µscans averaged and recorded per scan for the separate Fab chains. The EThcD active was set at true. The ions of interest were mass-selected by quadrupole in a 4 *m/z* isolation window and accumulated to the AGC target before fragmentation. MS/MS spectra were used to validate the sequences using LC-MS Spectator (Version 1.1.8313.28552) and ProSight Lite (1.4.8)[17,18]. In LC-MS Spectator, we used a tolerance of 10 ppm for MS1 and 20 ppm for MS2, respectively and applied the S/N threshold filtering (1.5). All the annotated ions were exported and visualized in ProSight Lite.

Native MS

To remove the *N*-glycan on the Fab, the samples were incubated with 1% Rapigest (Waters Corporation) for 3 min at 90 °C. PNGaseF was added to the sample and incubated at 50 °C for 10 min. Both the native Fab and the deglycosylated Fab were buffer exchanged into 150 mM ammonium acetate (pH 7.5) using Amicon 10 kDa MWCO centrifugal filters (Merck Millipore). The samples were loaded into gold-coated borosilicate capillaries (prepared in-house) and analyzed on a Q Exactive Ultra-High Mass Range (UHMR) Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectra were obtained in positive mode with an ESI voltage of 1.3 kV. The maximum injection time was set to 100 ms and the HCD energy was set to 100 V. The used resolution was 12,500 at 400 *m/z*. The

S-Lens level was set at 200. UniDec was used to generate the charge-deconvoluted spectrum[19].

Structural model of glycosylated MGUS Fab

The variable domain of the MGUS Fab was modelled using the ABodyBuilder2 webserver from the SAbPred suite. The predominant HexNAc(5)Hex(5)Fuc(1)NeuAc(2) glycoform was modelled as biantennary, bisected complex glycan with core fucosylation using the GLYCAM glycoprotein builder webserver. Figures were rendered in ChimeraX.

Results

Observation of a Fab-glycosylated IgG1 M protein

When analyzing the serum IgG1 clonal repertoire of a patient that had undergone a recent kidney transplant, we unexpectedly encountered an atypical antibody profile. This patient was part of a longitudinal study of patients who were observed after kidney transplantation for immunological monitoring. Per protocol in these patients, serum samples were obtained at different time points, starting from moments before surgery (t = 0 days) with follow-up samplings for over a year which were stored in a biobank. Samples of patients who developed designated bacterial infections in follow-up were drawn from the biobank and to investigate how the antibody repertoire responded to the surgery and subsequent infections, we applied our LC-MS IgG1 profiling approach to these longitudinal serum samples.

Strikingly, the IgG1 repertoire of this kidney transplant patient was very different from that of other donors, being dominated by seemingly a few extremely abundant clones that also exhibit relatively high masses (**Figure 1A**). This pattern remained unchanged before or after kidney transplantation or even after an episode of sepsis with a Klebsiella species (**Figure S1**).

Because these antibodies were detected at relatively high masses, we hypothesized that the unusual antibody profile of the patient resulted from a single IgG1 clone of very high concentration that may carry Fab glycosylation. Based on experimental data from our lab and theoretical calculations using sequences from the ImMunoGeneTics (IMGT) database[20], the bulk of IgG1 Fabs have backbone masses of roughly 46-50 kDa. The abundant Fabs that we detected, however, have higher masses of more than 50 kDa. This strongly suggests that they are modified by *N*-glycosylation, which would contribute roughly 2 kDa to their mass. Combining the signal intensities of the multiple putative glycoforms, the total concentration of this clone is approximately 10 mg/mL at t = 0, remaining high throughout the longitudinal follow-up (**Figure S1**). Compared to the total IgG1 concentration in human plasma, approximately 8 mg/mL on average, this is extremely high, prompting additional clinical tests. The patient was tested for an M protein using serum immunofixation, which confirmed the IgG kappa M protein and serum electrophoresis which could not quantify the M protein due to the low amount. In addition, free light chains (FLC) were determined (Binding Site®), demonstrating an elevated FLC kappa of 61.52 mg/L, and a FLC ratio of 3.92 (normal range 0.26-1.65). No bone marrow

biopsy was done and a diagnosis of monoclonal gammopathy of undetermined significance (MGUS) was made.

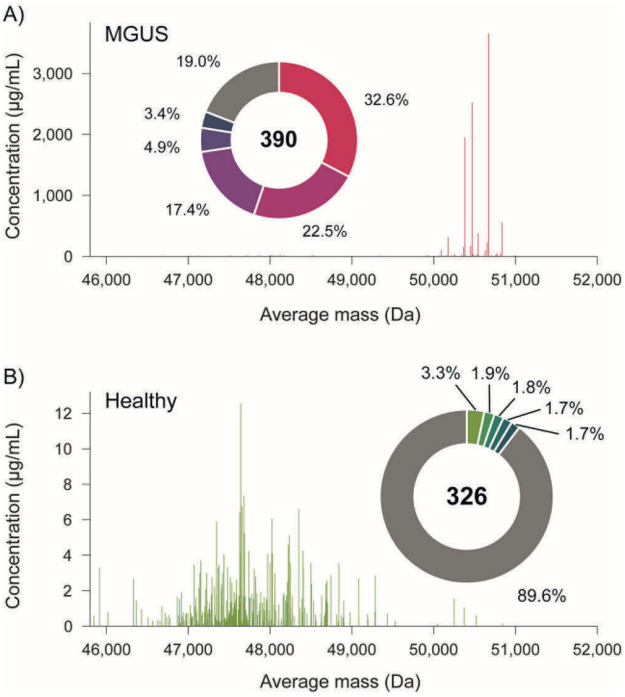


Figure 1 | Detection of MGUS by LC-MS-based IgG1 Fab profiling. (A) Serum IgG1 Fab profile of human subject with putative MGUS. **(B)** Illustrative IgG1 Fab profile from serum of a healthy human donor. Note the difference in complexity, average mass, and concentration of detected IgG1 Fabs, hinting at the presence of a Fab-glycosylated M protein in (A). Pie charts show the relative abundance (%) of the top 5 most abundant Fab identifications.

Direct MS-based sequencing and glycan localization of the serum-derived M protein

We have recently demonstrated the direct MS-based sequencing of serum-derived antibodies, using bottom-up proteomics methods[2]. Owing to the high abundance of the MGUS M protein in the serum samples of this donor, we were able to sequence the full antibody without further fractionation. First, we used an in-gel digestion protocol, using in parallel four proteases of complementary specificity, to obtain overlapping peptides for *de novo* sequencing by LC-MS/MS analysis. We identified the M protein as consisting of an IGHV4-4 heavy chain, coupled with an IGKV3-20 light chain. Notably, this experiment was performed with intact *N*-glycosylation and resulted in a lack of coverage in CDR1 of the heavy chain. The germline sequence of IGHV4-4 CDRH1 contains 5 serine residues and a single asparagine, priming it to obtain an *N*-glycosylation sequon by as many as 6 independent substitutions. These observations pointed to CDRH1 as a likely region to contain the putative Fab glycosylation.

Digestion with PNGase F results in the removal of the *N*-glycan and converts the glycan-linked asparagine to an aspartic acid residue[21]. We digested the serum sample with

PNGase F, followed by proteolysis with chymotrypsin and thermolysin in parallel. This recovered the previously missing CDRH1 sequence, containing a clear DSS motif, which would have corresponded with an NSS glycosylation sequon in the antibody before PNGase F digestion. Using the experimentally determined sequence, we then performed a glycoproteomics database search including common human *N*-glycans and were able to detect a predominant HexNAc(5)Hex(5)Fuc(1)NeuAc(2) glycan at the identified NSS sequon in CDRH1 (**Figure S2**).

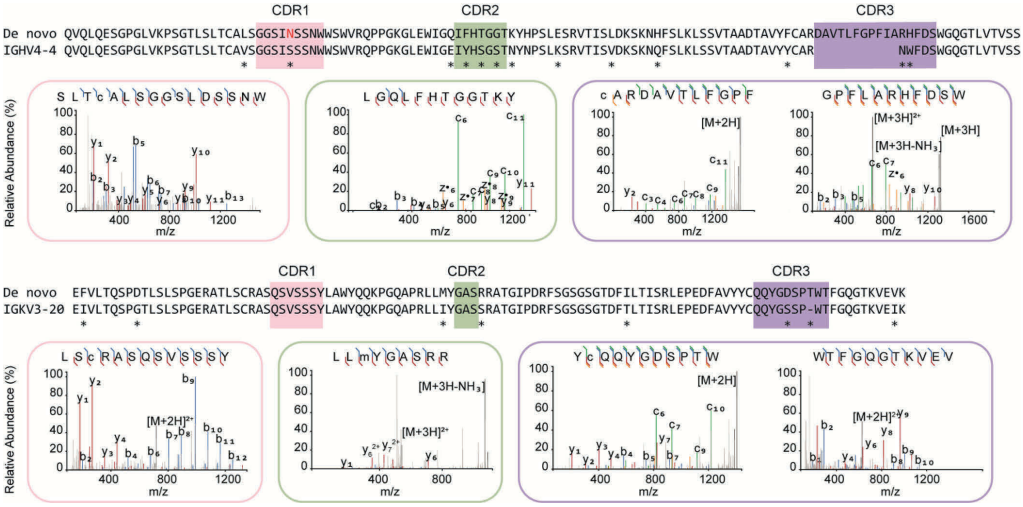


Figure 2 | *De novo* sequencing of the M protein Fab by bottom-up proteomics. The variable region alignment to the inferred germline sequence is shown for both heavy and light chains. Positions with putative somatic hypermutation are highlighted with asterisks (*). The MS/MS spectra supporting the annotation of the CDRs are shown beneath the sequence alignment. *b*/*y* ions are indicated in blue and red, while *c*/*z* ions are indicated in green and yellow.

Validation of the sequence and glycosylation of the Fab by middle-down LC-MS/MS

To validate our bottom-up *de novo* sequencing result, we further analyzed the MGUS Fab by native MS and middle-down LC-MS/MS. The intact mass profile of the Fab is shown in **Figure 3A**. The observed masses are consistent with the determined sequence, considering a pyroglutamic acid modification of the heavy chain *N*-terminus, and reveal heterogeneous glycosylation with a predominant HexNAc(5)Hex(5)Fuc(1)NeuAc(2) glycan, as also observed by bottom-up LC-MS/MS (**Table S1**). A predicted structural model of the variable domain shows this glycan protruding outwards from CDRH1, leaving the other CDR loops exposed (**Figure 3B**). The observed Fab glycosylation pattern follows a similar trend as reported by Bondt *et al.* in that it is enriched in galactosylation, sialylation and bisection compared to Fc glycosylation (**Figure 3C**; a full overview of glycoforms of the MGUS Fab is provided in **Table S1**)[22].

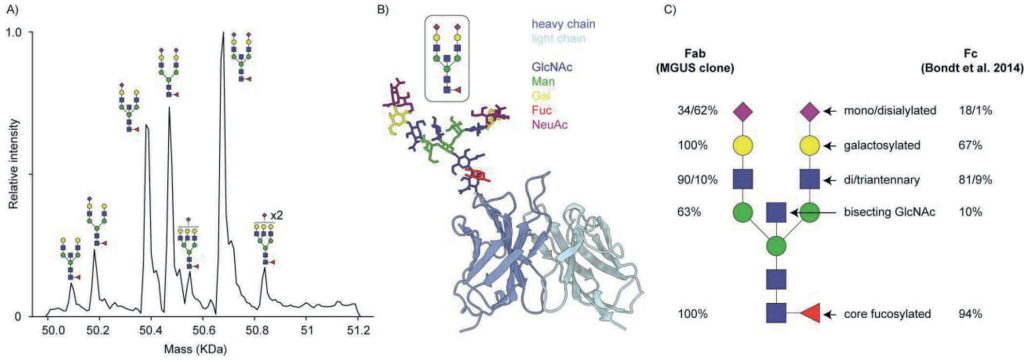


Figure 3 | *N*-Glycosylation of the M protein Fab. (A) Intact mass profile from native MS; peaks are annotated according to the assigned glycan structure. **(B)** ABodyBuilder2 structural model prediction of the MGUS Fab variable domain with HexNAc(5)Hex(5)Fuc(1)NeuAc(2) grafted on CDRH1 using GLYCAM. **(C)** Glycosylation profile of MGUS Fab compared to "normal" Fc glycosylation according to Bondt *et al.*[22].

We performed middle-down fragmentation of the reduced Fab by EThcD to confirm the sequence determined by bottom-up proteomics. This resulted in a coverage of 25.7% for the Fd and 60.5% for the LC (**Figure 4** and **S3**). We attribute the lower coverage of the Fd to the presence of the *N*-glycan in CDRH1, which is likely to fragment during EThcD, producing more complex spectra, with lower signals for individual fragments. Furthermore, glycan fragmentation is not yet implemented in currently available peak matching algorithms for middle-down LC-MS/MS. Nonetheless, the intact masses and middle-down fragmentation patterns support the M protein sequence determined by bottom-up proteomics.

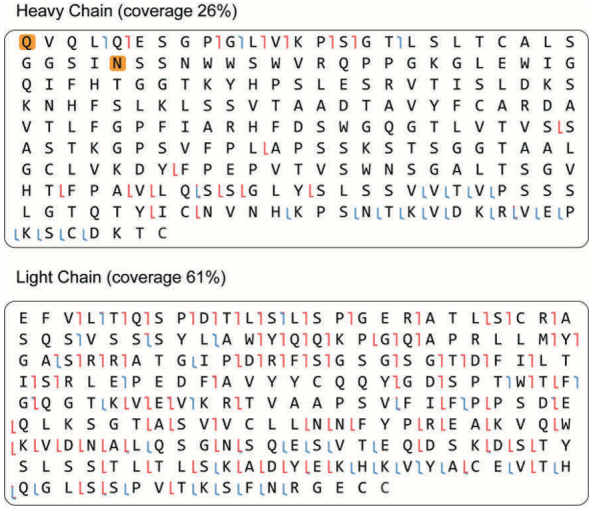


Figure 4 | Top-down LC-MS/MS data of the Fab-glycosylated M protein. Shown are the heavy and light chain sequences. Positions supported by *b*/*c* or *y*/*z* ions are indicated in red and blue, respectively.

Conclusion

Here we demonstrate that LC-MS-based IgG1 profiling of patient serum can lead to the detection of an M protein, which can be related to diseases such as MGUS or MM. The associated sequence of the M protein can be fully derived by mass spectrometry. In this case, mass spectrometry also revealed the presence, location and composition of Fab glycosylation in the heavy chain of the M protein. The outlined approach adds to an expanding mass spectrometry-based toolkit for characterizing monoclonal gammopathies such as MGUS and MM with fine molecular detail[23-28]. The ability to detect monoclonal gammopathies and determine M protein sequences straight from peripheral blood samples by mass spectrometry provides opportunities to understand the molecular mechanisms of these diseases.

Data Availability

The raw LC-MS/MS files and analyses have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD042266.

Acknowledgements

This research was funded by the Dutch Research Council NWO Gravitation 2013 BOO, Institute for Chemical Immunology (ICI; 024.002.009) to J.S and through the NWO TTW-NACTAR Grant #16442 (to A.J.R.H. and S.H.M.R.). We additionally acknowledge support from NWO, funding the MS facilities through the X-omics Road Map program (project 184.034.019).

References

1. Glavey, S.V. & Leung, N. Monoclonal gammopathy: The good, the bad and the ugly. *Blood Rev.* **30**, 223-231 (2016).
2. Dhodapkar, M.V. MGUS to myeloma: a mysterious gammopathy of underexplored significance. *Blood* **128**, 2599-2606 (2016).
3. Male, D., Stokes Peebles, R. & Male, V. in Immunology, Edn. 9th (Elsevier, 2020).
4. Dasari, S. et al. Detection of Plasma Cell Disorders by Mass Spectrometry: A Comprehensive Review of 19,523 Cases. *Mayo Clin. Proc.* **97**, 294-307 (2022).
5. Huhn, C., Selman, M.H.J., Ruhaak, L.R., Deelder, A.M. & Wuhrer, M. IgG glycosylation analysis. *Proteomics* **9**, 882-913 (2009).
6. Kinoshita, N. et al. Glycosylation at the Fab Portion of Myeloma Immunoglobulin G and Increased Fucosylated Biantennary Sugar Chains: Structural Analysis by High-Performance Liquid Chromatography and Antibody-Lectin Enzyme Immunoassay Using Lens culinaris Agglutinin1. *Cancer Res.* **51**, 5888-5892 (1991).
7. Mittermayr, S. et al. Polyclonal Immunoglobulin G N-Glycosylation in the Pathogenesis of Plasma Cell Disorders. *J. Proteome Res.* **16**, 748-762 (2017).
8. Bondt, A. et al. Human plasma IgG1 repertoires are simple, unique, and dynamic. *Cell Systems* **12**, 1131-1143.e1135 (2021).
9. Spoerry, C. et al. Novel IgG-Degrading Enzymes of the IgdE Protease Family Link Substrate Specificity to Host Tropism of Streptococcus Species. *PLoS One* **11**, e0164809 (2016).

10. Cassidy, J. & Nordby, G. Human serum immunoglobulin concentrations: Prevalence of immunoglobulin deficiencies+. *Journal of Allergy and Clinical Immunology* **55**, 35-48 (1975).
11. Gonzalez-Quintela, A. et al. Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clinical and Experimental Immunology* **151**, 42-50 (2008).
12. Tran, N.H., Zhang, X., Xin, L., Shan, B. & Li, M. De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci.* **114**, 8247-8252 (2017).
13. Tran, N.H. et al. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **16**, 63-66 (2019).
14. Schulte, D., Peng, W. & Snijder, J. Template-Based Assembly of Proteomic Short Reads For De Novo Antibody Sequencing and Repertoire Profiling. *Anal. Chem.* **94**, 10391-10399 (2022).
15. Brademan, D.R., Riley, N.M., Kwiecien, N.W. & Coon, J.J. Interactive Peptide Spectral Annotator: A Versatile Web-based Tool for Proteomic Applications. *Mol Cell Proteomics* **18**, S193-S201 (2019).
16. Bern, M., Kil, Y.J. & Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Current Protocols in Bioinformatics* **40**, 13.20.11-13.20.14 (2012).
17. Fellers, R.T. et al. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics* **15**, 1235-1238 (2015).
18. Park, J. et al. Informed-Proteomics: Open Source Software Package for Top-down Proteomics. *Nat. Methods* **14**, 909-914 (2017).
19. Marty, M.T. et al. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370-4376 (2015).
20. Lefranc, M.P. IMGT® databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, http://imgt.cines.fr. *Leukemia* **17**, 260-266 (2003).
21. Mann, A.C., Self, C.H. & Turner, G.A. A general method for the complete deglycosylation of a wide variety of serum glycoproteins using peptide-N-glycosidase-F. *Glycosylation & Disease* **1**, 253-261 (1994).
22. Bondt, A. et al. Immunoglobulin G (IgG) Fab Glycosylation Analysis Using a New Mass Spectrometric High-throughput Profiling Method Reveals Pregnancy-associated Changes *. *Mol. Cell. Proteomics* **13**, 3029-3039 (2014).
23. Barnidge, D.R. et al. Using Mass Spectrometry to Monitor Monoclonal Immunoglobulins in Patients with a Monoclonal Gammopathy. *J. Proteome Res.* **13**, 1419-1427 (2014).
24. Deighan, W.I. et al. Development of novel methods for non-canonical myeloma protein analysis with an innovative adaptation of immunofixation electrophoresis, native top-down mass spectrometry, and middle-down de novo sequencing. *Clinical Chemistry and Laboratory Medicine (CCLM)* **59**, 653-661 (2021).
25. Dupré, M. et al. De Novo Sequencing of Antibody Light Chain Proteoforms from Patients with Multiple Myeloma. *Anal. Chem.* **93**, 10627-10634 (2021).
26. McDonald, Z., Taylor, P., Liyasova, M., Liu, Q. & Ma, B. Mass Spectrometry Provides a Highly Sensitive Noninvasive Means of Sequencing and Tracking M-Protein in the Blood of Multiple Myeloma Patients. *J. Proteome Res.* **20**, 4176-4185 (2021).
27. Noori, S. et al. Retrospective Longitudinal Monitoring of Multiple Myeloma Patients by Mass Spectrometry Using Archived Serum Protein Electrophoresis Gels and De Novo Sequence Analysis. *HemaSphere* **6** (2022).
28. Noori, S. et al. Dynamic monitoring of myeloma minimal residual disease with targeted mass spectrometry. *Blood Cancer J.* **13**, 30 (2023).

Supporting Information

Supplementary Figures and Tables

Table S1 | Intact mass analysis of MGUS Fab and derivatives. Glycoforms are denoted with: N-acetyl hexosamine (N), hexose (H), fucose (F), and N-acetyl neuraminic acid (S).

Sample	Glycoform	MW_{exp} (Da)	SD (Da)[c]	MW_{theo} (Da)[d]	ΔMW (Da)	Relative intensity (%)
Intact Fab[a]	N5H5F1	50,093.96	0.40	50,090.79	3.17	3.63
	N4H5F1S1	50,181.62	0.16	50,178.85	2.77	7.52
	N5H5F1S1	50,384.96	0.21	50,382.05	2.91	22.04
	N4H5F1S2	50,472.65	0.02	50,470.11	2.54	23.78
	N5H6F1S1	50,549.69	0.04	50,544.19	5.50	4.93
	N5H5F1S2	50,675.96	0.13	50,673.30	2.66	32.57
	N5H6F1S2	50,837.75	0.17	50,835.45	2.30	5.54
Fab + PNGase F[a]	-	48,120.43	0.12	48,117.67	2.76	-
Fab + TCEP HC[b]	N5H5F1S2	26,952.69	0.37	26,951.09	1.60	-
Fab + TCEP LC[b]	-	23,729.78	0.37	23,731.31	-1.53	-

[a] *Native MS*

[b] *LC-MS*

[c] *Average and standard deviations are calculated across the series of charge states for native MS, across 4 replicate measurements for LC-MS.*

[d] *Average theoretical mass considering disulfide bond formation and pyroglytamic acid conversion of the heavy chain N-terminus.*

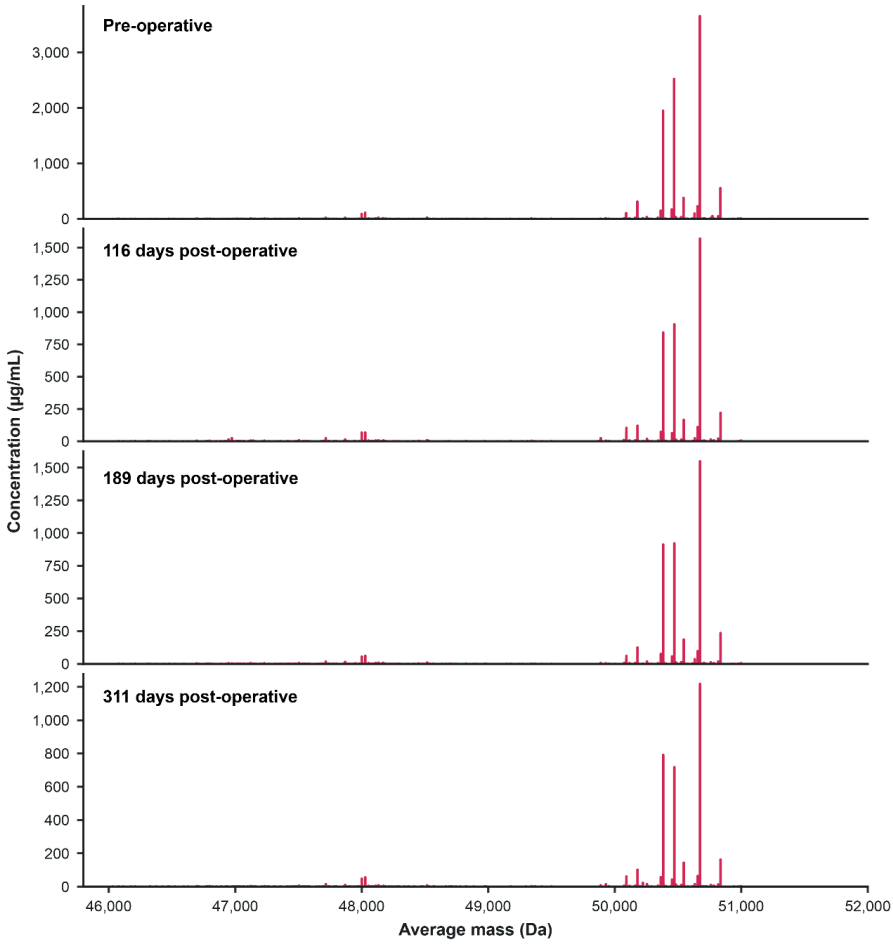


Figure S1 | Longitudinal IgG1 Fab profile of the MGUS patient.

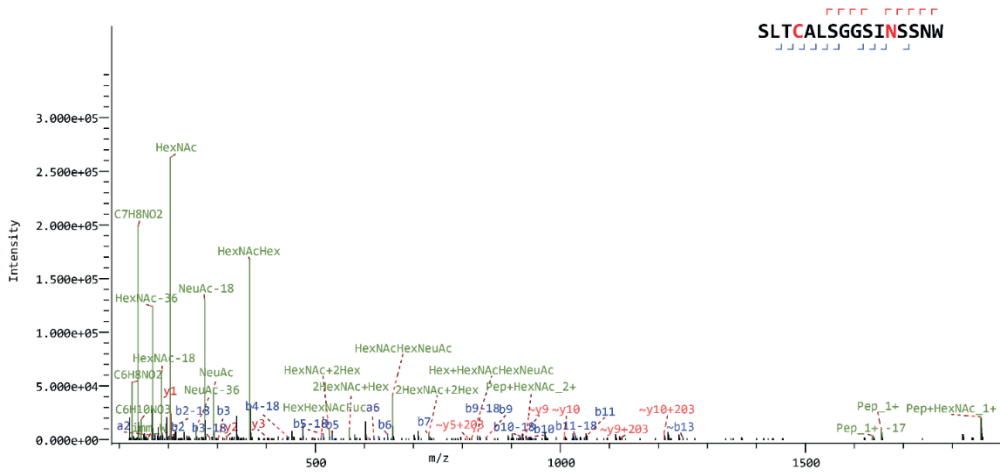


Figure S2 | Tandem MS spectrum of *N*-glycopeptide in CDRH1 of M-protein. Shown is the HexNAc(5)Hex(5)Fuc(1)NeuAc(2) glycoform.

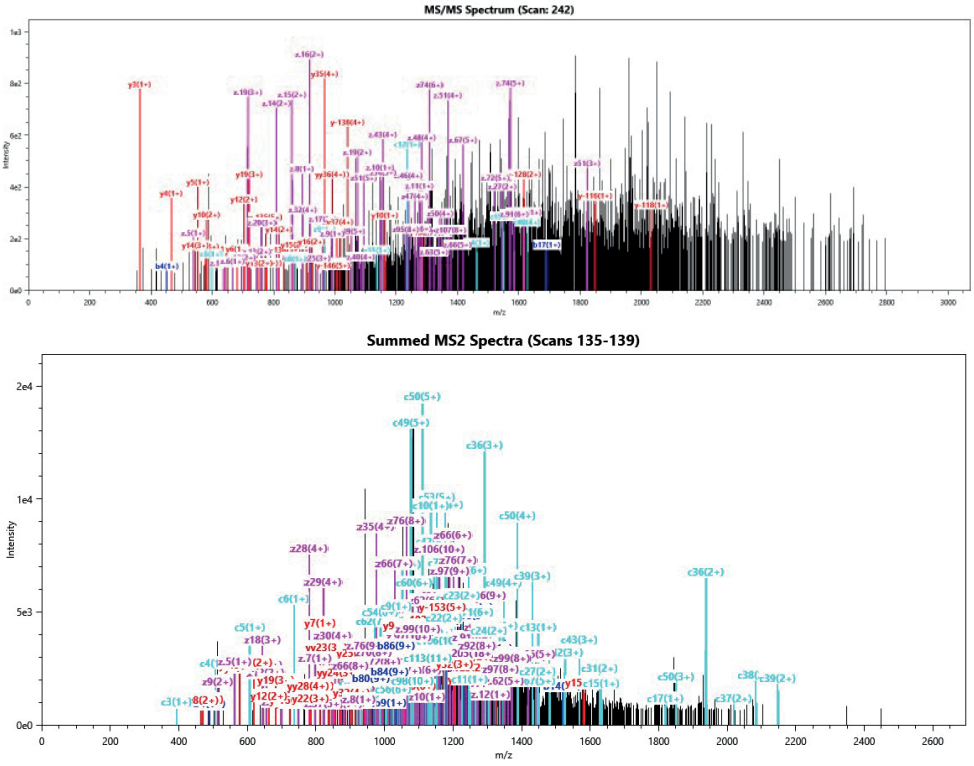
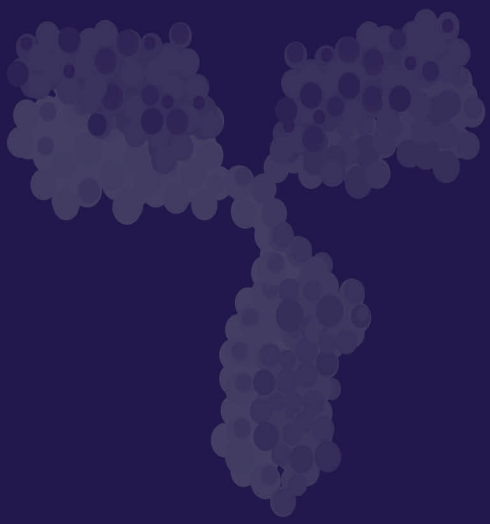


Figure S3 | Top-down LC-MS/MS spectra of M-protein Fab Heavy Chain (top) and Light Chain.

SYNOPSIS AND FUTURE PERSPECTIVES



Summary

This thesis focuses on the application of mass spectrometry to advance our understanding of antibody biology. Antibodies are a key component of our immune system's defense against harmful materials, and they are nowadays frequently used as biopharmaceuticals to specifically target agents involved in disease. Gaining better insight into how antibodies work and how they are generated therefore has great potential to improve modern medicine. In **Part 1** of my thesis, I contribute to our understanding of antibody structure and function by characterizing them through several mass-based analytical techniques. In **Part 2**, I develop mass-spectrometry-based tools for antibody repertoire analysis, which have the potential to generate a better insight into how they are generated by the body. This work includes the development of *de novo* sequencing techniques as a promising new tool for antibody sequence discovery. My findings and their relevance are summarized below.

Part 1: Investigating Antibody Biology through Mass Analysis

In **Chapter 1**, I introduce the basic concepts of antibody-mediated immunity. Also called immunoglobulins (Igs), these proteins are produced by the body to help recognize and clear pathogenic materials. This involves the generation of a highly variable part that specifically binds the target, which is connected to a constant part that interacts with receptors of the immune system. Antibodies exert various kinds of effector functions, of which particularly activation of the complement cascade plays a key role in the protection against bacteria and the activity of anti-tumor antibody therapies. Working towards a better understanding of antibody biology in both health and disease is thus of crucial importance in addressing the challenges of future medicine.

Chapter 2 introduces biomolecular mass spectrometry (MS) as an analytical technique to study proteins and their interactions, including those of antibodies. I describe the physical principles of mass spectrometry; the ionization of biomolecules and the different mass analyzers available to study them both qualitatively and quantitatively. Commonly, these instruments are applied to study whole proteomes in peptide-centric (or bottom-up) proteomics approaches. Native MS on the other hand analyses intact proteins with their fold and interactions remaining intact. This makes it possible to chart different protein isoforms and modifications, but also protein interactions. This includes the compositional analysis of protein complexes and the quantification of binding interactions.

In **Chapter 3**, I explore mass photometry (MP) as a relatively new analytical technique, benchmarking it against native (charge-detection, CD) MS and SEC-MALS for the analysis of antibodies and their interactions. While not rivalling native MS for mass resolution and accuracy in samples of limited complexity, this study shows that both MP and CDMS are valuable tools for studying antibody samples of high mass heterogeneity. In this context, MP arguably gives the most native overview of the sample content, whereas CDMS provides more accurate masses. Furthermore, while MP in its current form requires very low sample concentrations that may disrupt weak interactions, the technique can be used to quantify strong interactions of sub-μM affinities. However, in terms of quantitation,

SEC-MALS proved to be the most reliable technique. Combined, this shows that the three techniques are highly complementary, each occupying a different niche in the toolbox for studying antibodies and their interactions.

Chapter 4 applies native MS to study how Protein A of *Staphylococcus aureus* (SpA) enables these bacteria to evade antibody-mediated complement activation. We found that, even though SpA contains five binding domains, it can bind the constant segment of (human) IgG only 1:1, putatively forming two simultaneous interactions. In our experiments, this binding prevented IgGs from engaging in the Fc-Fc interactions that promote hexamerization on antigenic surfaces. As a result, recruitment of the six-armed complement recognition complex C1q was prevented, and therefore also activation of the cascade. Intriguingly, IgG3 was not affected, suggesting that formatting into this subclass could confer a major advantage to antibody therapeutics against this bacterium.

Chapter 5 is a showcase that demonstrates how a large toolbox of mass spectrometric approaches can be used to unravel new antibody biology. Combining MS with molecular biology techniques, this chapter redefines human circulatory IgM as a pentamer that universally contains CD5L as an additional structural component. Analyzing a cohort of healthy donors, serum proteomics first revealed a near-perfect correlation between the IgM and CD5L levels, with SEC-MS demonstrating that the two are associated. Next, a combination of proteomics, MP and CDMS showed that the canonical form of serum IgM is a pentamer with one copy of CD5L. Lastly, a comparison of IgM-CD5L to the previously used model showed a weaker interaction with several Fc receptor proteins, although the exact function of the CD5L component of IgM remains unknown.

Part 2: Towards Human Antibody Discovery by Mass Spectrometry

In **Chapter 6**, I introduce the different methods by which human antibodies may be discovered, focusing on the emerging role of MS in this process. Antibody sequences are generated by B lymphocytes when they are exposed to an antigen. When functional sequences have been established, these cells diverge into memory cells to store the initial sequence and plasma cells to produce antibodies. However, the number of memory sequences is orders of magnitude larger than that of effectively produced antibodies by plasma cells. While technical limitations have made memory B cells the primal source for antibody discovery, the most relevant unit of sequence information is thus the circulating antibody protein itself. Purified recombinant monoclonal antibodies can nowadays be reliably sequenced by peptide-centric (bottom-up) MS approaches, but for endogenous samples, complexity quickly becomes restrictive. This problem can be overcome by integrating bottom-up data with intact-level (top-down) analyses in one comprehensive approach. Combined with antibody repertoire screening techniques such as LC-MS clonal profiling, this makes MS a promising technique for future antibody discovery.

Chapter 7 explores the use of electron capture dissociation (ECD) for top-down fragmentation MS of the four human IgG subclasses. By analyzing the variable $F(ab')_2$ portion and removing the constant Fc, the signal could be focused on the parts that matter most. A combination of “pure” ECD – without additional activation – and the unique

disulfide linkages of antibodies ensured that a single type of fragment ion could be resolved. This resulted in easy-to-read ion ladders corresponding to sequences of the most crucial antigen-binding regions, the CDR3s. We found that the method worked well on all four IgG subclasses, although structural differences influenced the relative intensity and thus the detectability of certain fragments. Nonetheless, the method proved to provide crucial information that may become invaluable to complement peptide-centric approaches in *de novo* sequencing of endogenous antibodies.

Chapters 8 and **9** continue this line of work by applying similar ECD MS methods to IgA and IgM, extending it to highly heterogeneous and very large antibodies. While direct measurement of the intact antibodies resulted in very different and sometimes unresolved mass spectra, ECD fragment ion spectra were simple and highly similar. This was because the intrinsic heterogeneity of these antibodies is contained in their constant part, whereas ECD MS preferentially uncovered ion ladders from the variable part.

Lastly, **Chapter 10** brings together intact- and peptide-level MS analyses to discover and sequence an unexpected highly abundant patient antibody. Clonal profiling by LC-MS revealed that the IgG1 repertoire of a kidney transplant patient was dominated by a single Fab-glycosylated clone of extremely high concentration. This unexpected finding led to additional diagnostic tests and subsequent diagnosis of monoclonal gammopathy of undetermined significance (MGUS), a potential precursor to myeloma. The sequence of this antibody was fully revealed by combining top-down and bottom-up MS approaches, demonstrating the feasibility of this approach for discovering endogenous antibody sequences.

Nederlandse Samenvatting

Deze scriptie gaat over het inzetten van massaspectrometrie om onze kennis van menselijke antilichaambiologie te vergroten. Antilichamen zijn een essentieel onderdeel van ons immuunsysteem in de verdediging tegen schadelijke stoffen, en worden als zodanig steeds vaker toegepast als biofarmaceutisch middel om ziekten te behandelen. Betere kennis van hoe ze werken en hoe ze door het lichaam worden gegenereerd heeft daarom de potentie om een groot verschil te maken voor de toekomstige gezondheidszorg. In **Deel 1** van mijn scriptie beschrijf ik mijn onderzoek naar de structuren en functies van antilichamen, waarbij ik deze bestudeer met verschillende op massa gebaseerde analyse-technieken. In **Deel 2** beschrijf ik de ontwikkeling van massaspectrometriemethoden voor de analyse van antilichaamrepertoires, welke de potentie hebben om verdere inzichten te geven in hoe ons lichaam deze eiwitten genereert. Dit werk omvat het ontwikkelen van methoden voor *de novo* sequentiebepaling van antilichamen direct op eiwitniveau, een potentieel nieuwe manier van antilichaammedicijnontdekking. Mijn vindingen en hun relevantie vat ik hieronder samen.

Deel 1: Antilichaambiologie Bestudeerd doormiddel van Massa-analyse

In **Hoofdstuk 1** introduceer ik de basisconcepten van antilichaamgemedieerde immunologie. Antilichamen, ook wel immunoglobulinen (Ig) genoemd, zijn eiwitten die het lichaam helpen om pathogeen materiaal te herkennen en op te ruimen. Het maken van antilichamen begint met het genereren van een uniek variabel deel dat specifiek bindt aan het doelmateriaal, ofwel het antigeen. Dit antigeenherkennende deel is verbonden aan een constant deel dat interacties kan aangaan met receptoren van het immuunsysteem. Antilichamen kunnen verschillende effectorfuncties uitoefenen. Hierin is specifiek de activatie van het complementsysteem essentieel voor de bestrijding van bacteriën, en voor de werking van bepaalde antitumortherapieën. Kennis van hoe antilichaambiologie werkt tijdens gezondheid en ziekte is dan ook essentieel om uitdagingen in de gezondheidszorg beter het hoofd te kunnen bieden.

Hoofdstuk 2 introduceert biomoleculaire massaspectrometrie (MS) als analytische techniek om eiwitten en hun interacties te bestuderen, inclusief die van antilichamen. Dit hoofdstuk begint met een beschrijving van de fysische principes van massaspectrometrie, wat wil zeggen, de ionisatie van moleculen, en de verschillende massa-analysemethoden die beschikbaar zijn om deze moleculen kwalitatief en kwantitatief te bestuderen. Vaak wordt MS gebruikt om een proteoom – het totaal van eiwitten in een systeem – te bestuderen door het in kleine peptiden te knippen voor analyse met *bottom-up* proteomicsmethoden. Aan de andere kant van het spectrum bevindt zich natieve MS, waarmee gezuiverde eiwitten en eiwitcomplexen worden gemeten met een nog intacte structuur en de daaruit voortkomende interacties. Dit maakt het mogelijk om verschillende isovormen in kaart te brengen, om eiwitcomplexen compositioneel te analyseren, en om bindingsinteracties te kwantificeren.

In **Hoofdstuk 3** vergelijk ik massafotometrie (*mass photometry*, MP), een nieuwe analytische techniek, met gevestigde technieken zoals natieve MS en SEC-MALS voor de analyse van antilichamen en hun interacties. Hoewel MP vooralsnog niet kan tippen aan de massaresolutie en -accuraatheid van natieve MS, laat dit hoofdstuk zien dat zowel MP als ladingsdetectie (*charge-detection*, CD)MS zeer waardevol kunnen zijn voor de analyse van complexe antilichaammonsters. Hierin geeft MP wellicht het meest directe overzicht van alle deeltjes in het monster, terwijl CDMS de meest accurate massa's geeft. Bovendien bleek MP in staat om sterke antilichaaminteracties tot submicromolaire affiniteiten redelijk te kwantificeren. Wanneer het echter ging om nauwkeurig kwantificatie, bleef SEC-MALS de meest betrouwbare techniek. De drie technieken zijn dus zeer complementair aan elkaar, waarbij elk een unieke plek heeft in de gereedschapskist voor de analyse van antilichamen en hun interacties.

Hoofdstuk 4 behandelt de toepassing van natieve MS om te bestuderen hoe proteïne A van *Staphylococcus aureus* (SpA) deze bacteriën beschermt tegen antilichaam-gemedieerde activatie van het complementsysteem. Hoewel SpA vijf Ig-bindende domeinen heeft, vonden we dat dit eiwit 1:1 bindt aan menselijk IgG, met mogelijk twee simultane interacties. Deze binding van SpA bleek IgG-moleculen te verhinderen om Fc-Fc interacties aan te gaan, welke normaal gesproken op antigene oppervlakken plaatsvinden om IgG hexameren te vormen. Dit verhinderde de rekrutering van het zesarmige complementherkenningscomplex C1q, waardoor de activatie van het complementsysteem werd voorkomen. Van de vier IgG subklassen bleek IgG3 echter niet vatbaar te zijn voor SpA-binding, wat suggereert dat biofarmaceutische productie in deze subklasse voordelig kan zijn voor een antilichaamtherapie tegen *Staphylococcus aureus*.

Hoofdstuk 5 laat zien hoe de toepassing van een uitgebreide reeks aan massaspectrometriemethoden tot nieuwe ontdekkingen in de antilichaambiologie kan leiden. Door MS te combineren met moleculaire biologiemethoden, herdefinieert dit hoofdstuk de compositie van circulatoir IgM als een pentameer dat universeel gekoppeld is aan het CD5L-eiwit. Proteomicsexperimenten op serum van een cohort van gezonde donoren lieten een bijna perfecte correlatie tussen IgM- en CD5L-niveaus zien, terwijl SEC-MS aantoonde dat de twee eiwitten bovendien fysiek met elkaar geassocieerd zijn. Door een combinatie van proteomics, MP, CDMS en moleculairbiologische technieken toe te passen, konden we vervolgens bevestigen dat serum IgM een pentameer is met één CD5L molecuul dat covalent verbonden is via een disulfide. Tenslotte liet een vergelijking met CD5L-vrij IgM zien dat CD5L de interactie tussen IgM en enkele Fc-receptoren verzwakt, hoewel de exacte functie van CD5L-incorporatie in IgM nog onbekend blijft.

Deel 2: Richting Menselijke Antilichaamontdekking met Massaspectrometrie

In **Hoofdstuk 6** introduceer ik de verschillende methoden waarmee antilichaamsequenties kunnen worden ontdekt voor farmaceutische exploitatie, alsmede de opkomende rol van MS in dit proces. Antilichamen worden gegenereerd door B-lymfocyten wanneer ze worden blootgesteld aan een antigeen. Wanneer B-lymfocyten functionele antilichaamsequenties vinden, divergeren deze cellen in geheugencellen en plasmacellen die de sequentie bewaren en de corresponderende antilichamen produceren. Het aantal unieke geheugencellen is echter enorm veel groter dan het aantal uniek geproduceerde antilichamen. Technische beperkingen hebben er echter toe geleidt dat deze geheugencellen toch de meest gebruikte bron zijn geworden van farmaceutische antilichaamsequenties, ondanks het feit dat circulerende antilichaameiwitten de meest klinisch-relevante bron van antilichaamsequenties zijn. Voor zuivere monoclonale antilichamen kunnen deze eiwitsequenties tegenwoordig al betrouwbaar worden bepaald met peptide-niveau (*bottom-up*) proteomicsmethoden. Echter, wegens hun immense complexiteit is dit voor endogene antilichaammonsters al snel onmogelijk. De uitdaging kan niettemin worden aangegaan door peptidecentrische analyse te combineren met MS van intacte eiwitten in één geïntegreerde aanpak. In combinatie met antilichaamrepertoirescreening middels LC-MS, maakt dit MS een veelbelovende nieuwe techniek voor de ontdekking van functionele antilichaamsequenties.

Hoofdstuk 7 verkent het gebruik van elektronenvangstdissociatie (*electron capture dissociation*, ECD) voor intact-niveau fragmentatie van de vier IgG-subklassen tijdens MS-analyse. Door de constante Fc te verwijderen en de variabele $F(ab')_2$ te analyseren, konden we hier de focus van de analyse leggen op de delen van het molecuul die het belangrijkst zijn. Een combinatie van "pure" ECD zonder additionele activatie, en de unieke disulfidestructuur van de antilichamen, maakten het mogelijk dat één type fragment-ion duidelijk gescheiden kon worden van de rest. Hierdoor ontstonden makkelijk interpreteerbare ionenladders die de CDR3's, de meest cruciale antigeen-bindende sequenties, in kaart brachten. De methode werkte voor alle vier de IgG-subklassen, hoewel structurele verschillen invloed hadden op de intensiteit, en dus detecteerbaarheid, van de fragmenten. Desondanks leverde de methode cruciale informatie op die peptide-niveau analyses goed kan complementeren om de volledige *de novo* sequentiebepaling van endogene antilichamen te bereiken.

Hoofdstukken 8 en **9** zetten deze onderzoekslijn voort door vergelijkbare ECD MS methoden toe te passen op IgA en IgM. Deze antilichamen zijn veel heterogener en groter in massa dan IgG, wat MS-analyse bemoeilijkt. Hoewel directe meting van de intacte antilichamen tot sterk verschillende en soms zelfs onopgeloste massaspectra leidden, waren hun ECD-fragmentatiespectra relatief simpel en heel vergelijkbaar. Dit kwam doordat de intrinsieke heterogeniteit van deze antilichamen in hun constante deel zit, terwijl ECD MS preferentieel ionenladders vormt van het variabele deel.

Tenslotte brengt **Hoofdstuk 10** intact- en peptide-niveau MS samen om een abnormaal abundant antilichaam in patiëntmateriaal te detecteren en hiervan de sequentie te bepalen. LC-MS antilichaamrepertoire-analyse van een niertransplantatiepatiënt liet een

IgG1-profiel zien dat gedomineerd werd door één enkele Fab-geglycosyleerde kloon met een extreem hoge concentratie. Deze onverwachte vinding leidde tot additionele diagnostische tests en de diagnose van “monoklonale gammopathie van onbepaalde betekenis” (*monoclonal gammopathy of undetermined significance*, MGUS), een potentiële voorloper van myeloom. Door peptide- en intact-niveau MS te combineren, kon de volledige sequentie van dit antilichaam worden bepaald. Dit toont de haalbaarheid van deze aanpak aan voor het ontdekken van endogene antilichaamsequenties.

Discussion and Outlook

Our ability to characterize antibody responses is essential for understanding their roles in health and disease, and with that, our ability to create effective antibody therapies. In this respect, one of the holy grails in molecular biology is the ability to simultaneously identify and quantify all molecular entities and their interactions in a system. The field of biomolecular mass spectrometry has made great progress towards this aim, with different but complementary approaches solving different pieces of the puzzle. An ideal future technology would bridge these different levels of analysis to provide a comprehensive snapshot of all antibodies in a native biological system. Nonetheless, until this transformative technology is invented, many steps can be taken to improve what exists today.

Analyzing Antibody Interactions under Native Conditions

The study of antibody structures and mechanisms of action is largely comprised of the characterization of protein complexes and their interactions. Getting the most detailed information requires the study of antibodies under conditions as close as possible to their native environment in the body, as this will allow their interactions to remain intact and be captured. Developments in native MS technology have been instrumental in this context, with recent and ongoing advances extending capabilities even further.

In terms of nativeness, the ability to mass-analyze antibodies in standard biological buffering solutions or even more complex systems such as cell culture supernatants would be very advantageous. A key challenge to MS analysis, however, is the presence of non-volatile components such as salts. These are concentrated on the proteins during electrospray, forming non-specific adducts that convolute mass spectra. Soluble proteins are therefore generally buffer-exchanged to a volatile solution, which is time-consuming, not always possible, and a potential cause of artefacts and sample loss. One way to alleviate this obstacle would be to minimize the time spent in the new solution, for example through rapid online dilution or (chip-based) based separation. A preferred approach however improves the tolerance of MS itself to non-volatile components. This can be achieved by reducing spraying droplet size, as smaller droplets can hold fewer molecules to form non-specific adducts. Different approaches to this problem have already been pioneered, for example by preparing sub-micron diameter electrospray emitters that produce even smaller droplets[1,2]. While demonstrating proof-of-concept, reproducibly making these emitters out of traditionally used borosilicate has proved challenging and clogging seems to be a recurring problem. Nonetheless, further improvements may be realized by nanomaterial engineering, making native MS even more native by analyzing samples without buffer exchange.

Another key challenge in antibody research is the ability to assess their interactions with membrane proteins, not in the least because they often make attractive drug targets. Because membrane proteins are insoluble in aqueous solutions, this class of proteins remains markedly understudied compared to their soluble counterparts. This is because they require membranes or mimetics thereof to remain in a functional native-like state,

making them much more challenging to tackle analytically using traditional biophysical techniques – including native MS. In recent years, native MS has emerged as a rapid and highly sensitive technique to study membrane proteins and their interactions, almost becoming a separate discipline by itself[3,4]. Much effort has gone into the development of MS-compatible membrane mimetics to transfer proteins to the gas phase, as well as instrumental modifications that enable more extensive ion activation to strip away resulting adducts. The process remains very challenging, however, both in terms of maintaining native protein structure as well as in instrumental operation and data analysis for every single measurement. Future developments simplifying or automating this process could therefore be invaluable for the large-scale study of antibody interactions with membrane proteins.

Mass photometry (MP) through interferometric scattering microscopy (iSCAT) has proven to be highly complementary to native MS, both in the context of analyzing antibody interactions in standard buffering solutions and analyzing the interactions of membrane proteins[5,6]. MP's compatibility with salts and most biological buffering compounds makes it particularly suited for studying the interactions of antibodies with proteins that require high salinity or certain metal ions for their stability. Very recently, the technique has also been extended to membrane proteins, whereby these are solubilized in membrane mimetics and analyzed as a whole[6]. As an alternative approach, the measurement slides themselves could be coated with biological membrane systems. This would allow the study of interacting proteins, opening unique windows of opportunity not accessible by native MS.

While still no match in terms of mass accuracy and especially resolution, MP and iSCAT are still in their infancy and may experience future breakthroughs to narrow the gap. At the current stage, the technique suffers most from its intrinsically low signal-to-noise and indirect way of mass measurement. This limits resolution and the lower detection limit, as well as accurate mass measurement of compounds with undefined optical properties. However, future developments may alleviate some of these issues. While the study of antibody microheterogeneity and small molecule interactions may be out of reach, this could make the technique highly competitive for studying protein-protein interactions. This is especially true if also the dynamic range in sample concentration could be addressed, for example, through microfluidic rapid dilution or passivated measurement slides. Using alternative iSCAT approaches can also yield unique information not accessible through mass measurement alone[7]. As a microscopy technique, it also provides access to spatial information, which can be used to track proteins on biological membranes and even inside cells[8].

Combined, native MS and MP provide powerful tools for the characterization of antibodies and their interactions. As both technologies continue to improve, we can expect to see them yield even more detailed and accurate information about the structure and functions of antibodies. Additionally, synergetic combinations with other techniques, such as X-ray crystallography and cryo-electron microscopy, will provide an even more comprehensive understanding. Many biological questions remain, and both native MS and MP will play a vital role in advancing our understanding of how these important proteins function.

Towards Human Antibody Discovery by Mass Spectrometry

Advancing our understanding of how antibody responses are generated is crucial to our ability to confront diseases. In this context, it is especially of interest to chart antibody repertoires and how they change upon physiological events, as well as to determine the sequences of the constituent clones. Various techniques have been developed to gain insight into the diversity of antibody repertoires, starting with the study of immortalized B cells in the 1970s. Since then, enormous progress has been made, with single-cell genetic sequencing approaches nowadays able to process many thousands of B cells from individuals. Still, we are only scratching the surface of antibody repertoires, and current approaches are still limited by their indirect cell-based approach.

An ideal tool for repertoire analysis would identify and quantify all antibody molecules present, reveal their molecular composition, and provide information on their interactions. As demonstrated in this thesis, mass spectrometry-based approaches hold great promise in this respect, as they enable the analysis of antibodies directly on the protein level. This has several major advantages. Firstly, it guarantees that resulting sequences represent actively produced antibodies in their final most matured state. This is in contrast to B cell-based approaches, which also include many sequences that are less mature and may have never been selected for large-scale production. Secondly, protein-level analysis enables the quantitation and screening of antibody repertoires in their original polyclonal state, even in the absence of cell capture and *in vitro* antibody production. An MS-based analysis may thus provide more comprehensive information on functionalities, as well as synergies between multiple individual clones in the whole polyclonal response. The technique thus holds great promise, and with the help of further technological developments, it may soon be established as a new tool for the in-depth characterization of antibody repertoires.

Advancing Antibody Repertoire Analysis through Clonal Profiling

The first component of the MS toolbox for antibody repertoire analysis is clonal profiling, which identifies and quantifies clones independent of sequence information. We have now published methods that make this possible for IgG1 and IgA1 by analyzing their Fab molecules, which we obtained through specific enzymatic digestion. Surpassing our expectations when I started in the lab, we are now able to sometimes chart close to a thousand of the most abundant clones in a sample, explaining 50-95% of the total Ig concentration. This approach has enabled us to study the dynamics of antibody repertoires in response to physiological changes and to discern polyclonal responses generated upon infection.

While our IgG1 and IgA1 clonal profiling methods cover the most abundant subclasses of the two most abundant human isotypes, an expansion to other isotypes would be very useful. From an antibody biology point of view, IgM is especially relevant, as this is the first isotype generated in an immune response. Despite the remaining challenges in obtaining monodisperse Fab molecules from heterogeneously glycosylated IgM, we recently developed an isotype-specific light chain profiling method. While the lack of heavy chain

information makes the approach less suitable for antibody sequence discovery, it works very well as a proxy for assessing clonality. The approach enabled us to assess clonal overlap between the serum and milk IgM repertoires of individual donors, revealing that the two compartments are largely unique (**Figure 1**). By applying the same methodology also to different antibody isotypes, it may be possible to track the development of an antibody response through maturation and class switching, which could provide invaluable insights into the development of antibody responses.

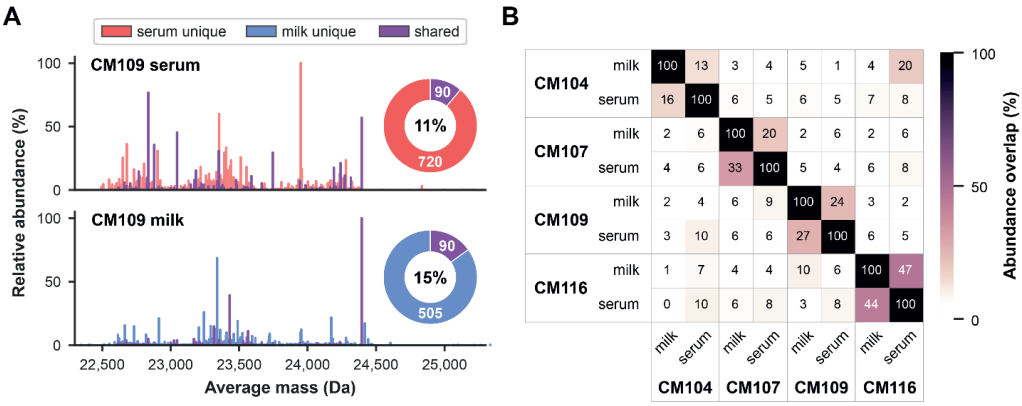


Figure 1 | IgM-specific LC profiling of serum and milk obtained from individual donors reveals two distinct repertoires. (A) Exemplary LC clonal profiles from plasma and milk IgM of the same donor. Red and blue peaks indicate clones unique to plasma and milk respectively, with purple peaks representing clones present in both. The total number of identifications is indicated in a donut chart. **(B)** Heatmap showing the relative abundance of matching IgM LCs between samples. While roughly 30% of the abundance in each sample constituted LC identifications also made in the complementary biofluid of the same donor, overlap between different donors was very low.

For the analysis of mature antibody repertoires against specific targets, an expansion to the remaining IgG subclasses (*i.e.*, IgG2, IgG3 and IgG4) is also highly relevant. Especially IgG2 is of interest as this subclass is thought to be crucial to responses against carbohydrate structures on bacterial and viral surfaces. While our current approaches work best for antibody Fab molecules, to our knowledge, currently available capturing materials and Fab-producing enzymes do not provide the required specificity for IgG2. Nonetheless, we have recently developed an alternative approach to analyzing total IgG on the level of F(ab')$_2$ molecules using an enzyme that cleaves IgG below the hinge. At an average mass of about 100 kDa – over twice the size Fabs – F(ab')$_2$ molecules are harder to analyze effectively by MS. However, by extensively tuning the instrumental parameters, we managed to improve sensitivity for these larger molecules. This enabled us to perform clonal profiling not only on IgG1 but also on the remaining IgG subclasses, from which we could identify in the order of 300 additional clones (**Figure 2**).

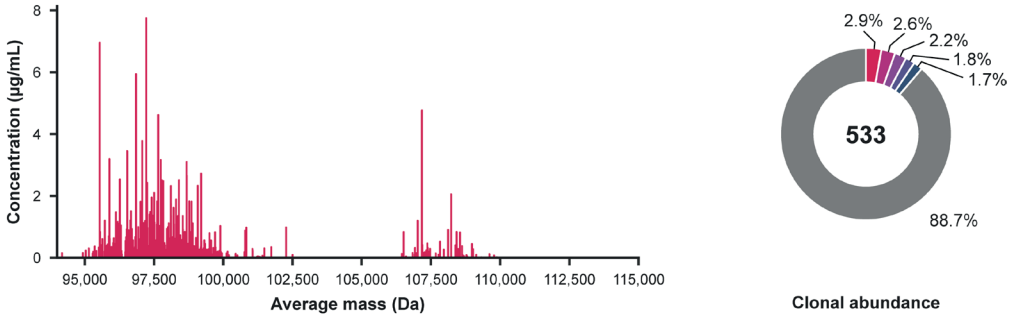


Figure 2 | Exemplary IgG1-4 clonal profile of a healthy donor. By capturing total IgG and cleaving the antibodies enzymatically below the hinge, we produced F(ab')$_2$ molecules that could be measured by intact LC-MS. In the clonal profile (left), each peak represents the average mass and concentration of a unique identified species. The distribution between 95 kDa and 105 kDa constitutes IgG1, IgG2 and IgG4 molecules. The secondary distribution with average masses over 105 kDa constitutes IgG3 molecules, which have an extended hinge region with *O*-glycosylation sites. In this sample, we identified and quantified a total of 533 unique species, in which the top 5 make up 11% of the total detected abundance (right).

While existing clonal profiling workflows can already provide us with a wealth of information, what they identify likely only represents the tip of the iceberg. Theoretical calculations have shown that the number of active clones may still be orders of magnitude larger than what we can identify experimentally[9]. The bulk of antibody repertoires would thus constitute lowly abundant clones, which require more effort to resolve or detect. Indeed, diving into the depth of the repertoire through target-specific profiling, we have now found that many additional lowly abundant clones can be identified when the signal is concentrated and distributed over an enriched fraction of clones. Excitingly, this means that much more can still be realized. If we can improve sensitivity, dynamic range, and resolution, it may thus soon be possible to chart not hundreds but perhaps thousands of clones by LC-MS.

Sensitivity and dynamic range are mostly engineering problems, of which the first is relatively straightforward to solve. Mass spectrometry instrumentation is optimized and tuned for the transmission of ions within a certain mass range. For most biological applications, this range is selected for the detection of small molecules and even smaller fragments thereof ($m/z \approx 500$). On the other end of the spectrum, intact protein complexes are the bread and butter of native MS instruments ($m/z \approx 8000$). Sensitivity for clonal profiling could thus be improved by using ion optics that are optimized for the transmission of denatured proteins ($m/z \approx 2000$), which can be done by extensive tuning of existing instruments or the development of dedicated top-down MS instrumentation. Another straightforward improvement would be to move from microflow to nanoflow chromatography, which further boosts the MS signal and reduces sample consumption. The dynamic range of the mass spectrometer is a more challenging problem, however, especially for Orbitrap instruments that impose strict limits both on the number of ions per scan and the ratio between the least and most abundant species. Intrascan dynamic range improves with microscans and higher resolution settings, though at the expense of

sensitivity and duty cycle length. Alternatively, time-of-flight instruments with a higher dynamic range could be used instead.

In terms of resolution – the ability to distinguish between clones – several steps can still be taken. On the MS side, one way to achieve this could be through increasing the instrumental resolution to an isotopic level, though the resulting longer duty cycle on current Orbitrap instruments quickly becomes prohibitive. Alternatively, charge-reduction or charge-detection methods could be used to deconvolute mass spectra. On the chromatography side, enhancing separation to sharpen elution peaks would lower the number of unique clones injected into the MS at the same time. Furthermore, orthogonal separation techniques could be used for a similar effect. This could include separation by high pH reversed phase, ion exchange, or online ion mobility chromatography.

Towards High-Throughput MS-Based Antibody Sequencing

In an ideal MS-based antibody sequencing approach, all information would be obtained directly from a single LC-MS clonal profiling experiment. Such a direct top-down approach would theoretically result in the highest sequencing signal-to-noise while negating the need for computationally challenging sequence assembly from peptide-level reads. The reason for this is that, in top-down MS, all relevant information is contained within a single precursor ion, which can be isolated with very high precision. The technology available today, however, cannot yield complete and predictable fragmentation of intact proteins just yet. The approach would thus require MS^n fragmentation, whereby the fragments themselves are isolated and fragmented again to reveal their amino acid sequence. While this may become feasible on an LC-MS timescale in the future, other complementary sequencing techniques already provide opportunities for improving antibody sequencing throughput today.

Bottom-up MS experiments nowadays reliably generate sequence reads that fully cover reasonably purified antibodies. Key challenges here are related to the assembly of these sequence reads, which becomes increasingly difficult with sample complexity. Compared to more matured genomic sequencing techniques, bottom-up MS yields a low sequence coverage depth. This is because peptides are not formed randomly, but by proteolytic cleavage at specific sites, which limits their length and overlap. While this problem is partially overcome by performing separate digestions with different proteases, alternative approaches to a more random generation of longer peptides could be beneficial. Alternatively, rapid developments in the field of nanopore peptide sequencing may provide a highly complementary platform in the future[10]. Nonetheless, when dealing with endogenous antibodies, purification to the level required for a pure peptide-based approach is only feasible for dominant and well-separated clones.

In the short term, *de novo* sequencing of endogenous antibodies will depend on the ability to combine multiple layers of MS analysis successfully and efficiently. Our current solution uses intact chain and fragment masses from top-down experiments as restraints for a template-based assembly of bottom-up sequence reads. While still a laborious process, there are many low-hanging fruits in terms of automation, both on the experimental and

data analysis side. LC-MS clonal profiling could be combined with simultaneous fraction collection, providing excellent starting materials for subsequent top-down and bottom-up characterization. Such fractions could be further processed using commercially available liquid handling platforms, followed by measurement and automated data analysis. Combined, this could already make it feasible to obtain sequences of tens to perhaps hundreds of clones in a sample, bringing the throughput of MS-based analysis a big leap forward.

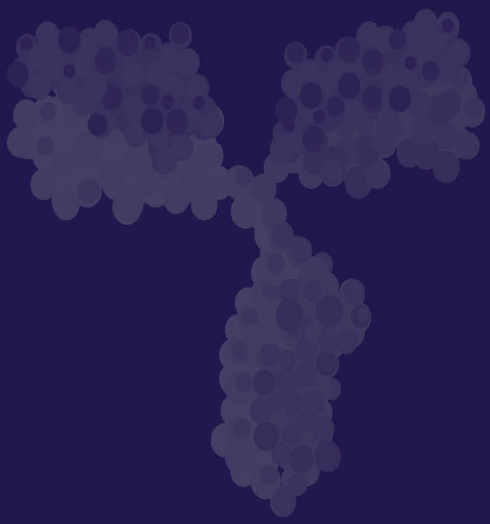
An Emerging Decade of MS-Based Antibody Discovery?

The field of antibody repertoire analysis and sequence discovery has taken a huge flight since the isolation and recombinant production of the first clones in the 1970s. Obtained using nowadays rudimentary cloning techniques, the first therapeutic antibody was registered in 1986. The availability of practical DNA sequencing methods and phage display screening led to a first boom starting in the 1990s, followed by a second spurt in the 2010s with the availability of (single-cell) high-throughput next-generation sequencing techniques. While all technologies until today have identified antibodies indirectly through B cells, I summarized here how mass spectrometry can provide unique advantages through direct protein-level analysis. With technology rapidly developing, perhaps the 2020s will be the decade where MS-based sequencing is established to bring antibody therapeutics to the next level.

References

1. Jordan, J.S., Xia, Z. & Williams, E.R. Tips on Making Tiny Tips: Secrets to Submicron Nanoelectrospray Emitters. *J. Am. Soc. Mass. Spectrom.* **33**, 607-611 (2022).
2. Chen, S., Zeng, J., Zhang, Z., Xu, B. & Zhang, B. Recent advancements in nanoelectrospray ionization interface and coupled devices. *Journal of Chromatography Open* **2**, 100064 (2022).
3. Bolla, J.R., Agasid, M.T., Mehmood, S. & Robinson, C.V. Membrane Protein–Lipid Interactions Probed Using Mass Spectrometry. *Annual Review of Biochemistry* **88**, 85-111 (2019).
4. Robinson, C.V. Mass spectrometry: From plasma proteins to mitochondrial membranes. *Proc. Natl. Acad. Sci.* **116**, 2814-2820 (2019).
5. Asor, R. & Kukura, P. Characterising biomolecular interactions and dynamics with mass photometry. *Curr. Opin. Chem. Biol.* **68**, 102132 (2022).
6. Olerinyova, A. et al. Mass Photometry of Membrane Proteins. *Chem* **7**, 224-236 (2021).
7. Taylor, R.W. & Sandoghdar, V. Interferometric Scattering Microscopy: Seeing Single Nanoparticles and Molecules via Rayleigh Scattering. *Nano Lett.* **19**, 4827-4835 (2019).
8. Sandoghdar, V., Kueppers, M., Albrecht, D., Kashkanova, A. & Luehr, J. (Research Square, 2022).
9. Lavinder, J.J., Horton, A.P., Georgiou, G. & Ippolito, G.C. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Curr. Opin. Chem. Biol.* **24**, 112-120 (2015).
10. Meller, A. Are nanopore technologies ready for the proteomic challenge primetime? *Mol. Cell* **82**, 237-238 (2022).

APPENDIX



Acknowledgements

Writing this thesis would not have been possible without the help of some exceptional people. Here, I would like to take the opportunity to say thanks to all those that have helped me along the way.

First and foremost, I want to extend my gratitude to my promoter **Albert** for giving me the amazing opportunity to take on this PhD project. I appreciated your hands-off approach and trust in me, which enabled me to take ownership of my work and grow quickly. Additionally, I am also very grateful for the support you gave me in pursuing my dream of translating our academic work into a startup company. We are finally co-founding Abvion, and I feel fortunate to have you as a partner in this new adventure.

This work would also not have been possible without the help of all my colleagues in the **BioMS group**. Thank you all for providing me with such a great environment over the last few years. I have gotten to know many of you as talented and devoted scientists, who furthermore are always willing to extend a helping hand. There are too many of you to thank personally, but there are a few of you I want to give special mention to below.

First, I want to give a big thank you to **Arjan**, probably my biggest supporter in the lab both professionally and as a friend. You have given me the best MS training I could have asked for and were always eager to share your expertise by helping me with my projects and my (sometimes sandpaper-involving) fabrications. Equally as important is your sense of humor, bringing countless moments of laughter and helping me to keep going strong.

Next, I want to thank the people I have worked most closely with. First up is **Albert B**, my office mate, paranymph, and favorite immunoglobuloligist (*i.e.*, an esteemed expert in "immunoglobulomics"). You are a talented and knowledgeable researcher, and I am grateful for all the help and input you have given me over the years. I also want to thank you for bringing me into your initiative to valorize our research, which ultimately formed the basis for founding Abvion. Thank you, **Jean-Francois**. Your work has been instrumental for many chapters of this thesis. I have gotten to know you as a brilliant and diligent mass spectrometrist, and I learned a lot from working with you. **Sem**, working with you was always a pleasure, and it led to some great publications for both of us. Thank you for your teamwork and I hope our paths will cross again in the future. I also want to thank team IgM, including **Tim**, **Szu-Hsueh**, **Marie**, **Victor**, **Amber**, **Karli**, **Marta**, **Mirjam**, **Arjan**, and **Albert B**. Thank you **Weiwei** and **Joost** for picking up the MGUS sequencing project. Weiwei, also thank you for being very cordial with me and for training me in your bottom-up antibody sequencing (and cake-baking) workflows. Lastly, I want to give a special mention to **Douwe** and **Bastiaan** for helping me with my questions about coding and data analysis for de novo sequencing, and to **Tess** for the good work she did during her internship with Jean-Francois and me.

My work in the BioMS group also would not have been possible without the help of the support staff. Thank you all very much, especially **Mirjam**, **Arjan**, **Nadia**, **Naomi,** and **Dominique** for your technical support and training. I also want to thank **Corine** for making everything run smoothly on the administrative side and for initiating many memorable social events.

Over the years, I have also had the privilege of being part of many fruitful collaborations. Firstly, **Suzan**, thank you for the great collaboration with the UMCU and your generosity over the years. Working with you and your lab has boosted my PhD, both in terms of output and in improving my knowledge about immunology. Especially also **Sjors** and **Rita** have been instrumental in this, so thank you for your team spirit and for the great work we have done together. I also want to thank **Carla** and **Remy** for working together on IgM, and Carla again for the many antibody samples you made for me and my colleagues. Next, I want to thank all the great people from Genmab who I have had the pleasure of working with. Thank you, especially **Janine**, **Frank**, **Renoud**, **Xiaoguang**, **Muriel**, **Boris, Daphne,** and **Aran** for the fantastic collaboration and your support. Lastly, I also want to thank **Nienke**, **Theo,** and **Gestur** at Sanquin. It was an incredible coincidence that brought us together after making the same discovery and combining our forces allowed us to make more impact than we could have on our own.

I have been fortunate to have made several trips during my time in Utrecht, all of which ended up being unique experiences. First up was the #hybridroadtripbremenboys adventure with **Arjan** and **Jean-Francois** to the Thermo Fisher factory in Bremen. Although I could not contribute much, I had a great time. Three years later, the ASMS conference in Minneapolis was next, and I want to thank especially **Victor** and **Kelly G** for their great company. Finally, I want to thank **Danique** and **Tatiana** for our unforgettable trip to San Jose and San Francisco to visit the Thermo Fisher factory there. I could not have asked for two better travel mates for this adventure (and for watching Indian Matchmaking with).

I also want to thank everyone who helped me stay motivated and who made my experience in Utrecht much better, with a few special mentions. Firstly, thank you to my paranymph **Danique** for always being kind and supportive, both personally and professionally. Your cheerfulness and sense of humor always brighten up the room. Next, I want to thank my office mates **Albert B**, **Max**, **Domenico,** and **Eduard**. Max, thank you for your company and for always being willing to help me with my Python questions. Thank you also Eduard for being a great office mate and for your contagious sense of humor. I also want to thank **Julia**, who started with me in Utrecht on the same day, as well as **Rien** and **Linus** (a.k.a. Rinus), who are getting started just as I am going away.

Pursuing my ambitions of founding a company, I also had the privilege of working with many people who supported me. I want to thank my teammates in the Utrecht Inc validation program, **Albert B** and **Sem**, and also my startup support team, **Paul**, **Ton**, **Janine**, and **Erik** for their invaluable help.

Lastly, I want to thank my mom and dad, **Ellen** and **Lex**, as well as my brother **Frederik**. Whether it was navigating housing challenges in Utrecht or lending a sympathetic ear to my problems and frustrations, your support and guidance have been instrumental in shaping where I am today.

About the Author

Maurits den Boer was born on July 5th, 1991, in Raamsdonk, the Netherlands. After completing high school at Stedelijk Gymnasium Leiden, he discovered his fascination with biotechnology while spending a year to rekindle his interest in natural sciences. This led him to pursue a bachelor's degree in Life Science and Technology, a joint program offered by Delft University of Technology and Leiden University.

During his undergraduate studies, Maurits performed an internship in molecular tumor pathology under the guidance of prof. Johannes Morreau at the Leiden University Medical Center. Continuing his education at Leiden University, Maurits earned a master's degree (cum laude) in Life Science and Technology, specializing in Science-Based Business Development. This comprehensive program encompassed a research internship in protein chemistry, working with prof. Marcellus Ubbink to study proteins by NMR at the Leiden Institute of Chemistry. Additionally, he gained practical business knowledge through an internship at the venture capital fund of InnovationQuarter in The Hague.

In September of 2018, Maurits started his doctoral studies in the Biomolecular Mass Spectrometry and Proteomics group at Utrecht University under the supervision of prof. Albert Heck and prof. Suzan Rooijakkers. His research primarily focused on developing mass spectrometry-based tools for characterizing humoral immune responses and *de novo* sequencing of antibodies. The results of these doctoral studies have been published in multiple scientific journals, forming the foundation of this thesis on ***Exploring Humoral Immune Responses by Mass Spectrometry***.

List of Publications

Publications Included in This Thesis

1. **den Boer, M.A.***, Greisch, J.-F.*, Tamara, S.*, Bondt, A. & Heck, A.J.R. Selectivity over coverage in de novo sequencing of IgGs. *Chemical Science* **11**, 11886-11896 (2020).
2. Greisch, J.-F.*, **den Boer, M.A.***, Beurskens, F., Schuurman, J., Tamara, S., Bondt, A. & Heck, A.J.R. Generating Informative Sequence Tags from Antigen-Binding Regions of Heavily Glycosylated IgA1 Antibodies by Native Top-Down Electron Capture Dissociation. *J. Am. Soc. Mass. Spectrom.* **32**, 1326-1335 (2021).
3. Cruz, A.R.*, **den Boer, M.A.***, Strasser, J., Zwarthoff, S.A., Beurskens, F.J., de Haas, C.J.C., Aerts, P.C., Wang, G., de Jong, R.N., Bagnoli, F., van Strijp, J.A.G., van Kessel, K.P.M., Schuurman, J., Preiner, J., Heck, A.J.R. & Rooijakkers, S.H.M. Staphylococcal protein A inhibits complement activation by interfering with IgG hexamer formation. *Proc. Natl. Acad. Sci.* **118**, e2016772118 (2021).
4. Tamara, S.*, **den Boer, M.A.*** & Heck, A.J.R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* (2021).
5. Greisch, J.-F., **den Boer, M.A.**, Lai, S.-H., Gallagher, K., Bondt, A., Commandeur, J. & Heck, A.J.R. Extending native Top-Down electron capture dissociation to MDa immunoglobulin complexes provides useful sequence tags covering their critical variable Complementarity-Determining regions. *Anal. Chem.* **93**, 16068-16075 (2021).
6. **den Boer, M.A.**, Lai, S.-H., Xue, X., van Kampen, M.D., Bleijlevens, B. & Heck, A.J.R. Comparative analysis of antibodies and heavily glycosylated macromolecular immune complexes by size-exclusion chromatography multi-angle light scattering, native charge detection mass spectrometry, and mass photometry. *Anal. Chem.* **94**, 892-900 (2021).
7. Oskam, N.*, **den Boer, M.A.***, Lukassen, M.V., Ooijevaar-de Heer, P., Veth, T.S., van Mierlo, G., Lai, S.-H., Derksen, N.I.L., Yin, V.C., Streutker, M., Franc, V., Siborova, M., Damen, M., Kos, D., Barendregt, A., Bondt, A., van Goudoever, J.B., de Haas, C.J.C., Aerts, P.C., Muts, R.M., Rooijakkers, S.H.M., Vidarsson, G., Rispens, T. & Heck, A.J.R. CD5L is a canonical component of circulatory IgM. *bioRxiv*, 2023.2005.2027.542462 (2023).
8. Peng, W., **den Boer, M.A.**, Tamara, S., Mokiem, N.J., van der Lans, S.P.A., Bondt, A., Schulte, D., Haas, P.-J., Minnema, M.C., Rooijakkers, S.H.M., van Zuilen, A.D., Heck, A.J.R. & Snijder, J. Direct Mass Spectrometry-Based Detection and Antibody Sequencing of Monoclonal Gammopathy of Undetermined Significance from Patient Serum: A Case Study. *J. Proteome Res.* (2023).

* *Equally contributing authors*

Other Publications

9. van Keulen, B.J., Romijn, M., Bondt, A., Dingess, K.A., Kontopodi, E., van der Straten, K., **den Boer, M.A.**, Burger, J.A., Poniman, M., Bosch, B.J., Brouwer, P.J.M., de Groot, C.J.M., Hoek, M., Li, W., Pajkrt, D., Sanders, R.W., Schoonderwoerd, A., Tamara, S., Timmermans, R.A.H., Vidarsson, G., Stittelaar, K.J., Rispens, T.T., Hettinga, K.A., van Gils, M.J., Heck, A.J.R. & van Goudoever, J.B. Human Milk from Previously COVID-19-Infected Mothers: The Effect of Pasteurization on Specific Antibodies and Neutralization Capacity. *Nutrients* **13**, 1645 (2021).
10. Zwarthoff, S.A., Widmer, K., Kuipers, A., Strasser, J., Ruyken, M., Aerts, P.C., de Haas, C.J.C., Ugurlar, D., **den Boer, M.A.**, Vidarsson, G., van Strijp, J.A.G., Gros, P., Parren, P.W.H.I., van Kessel, K.P.M., Preiner, J., Beurskens, F.J., Schuurman, J., Ricklin, D. & Rooijakkers, S.H.M. C1q binding to surface-bound IgG is stabilized by $C1r_2s_2$ proteases. *Proc. Natl. Acad. Sci.* **118**, e2102787118 (2021).
11. Bondt, A., Hoek, M., Tamara, S., de Graaf, B., Peng, W., Schulte, D., van Rijswijck, D.M.H., **den Boer, M.A.**, Greisch, J.-F., Varkila, M.R.J., Snijder, J., Cremer, O.L., Bonten, M.J.M. & Heck, A.J.R. Human plasma IgG1 repertoires are simple, unique, and dynamic. *Cell Systems* **12**, 1131-1143.e1135 (2021).

12. Chataigner, L.M.P., Gogou, C., **den Boer, M.A.**, Frias, C.P., Thies-Weesie, D.M.E., Granneman, J.C.M., Heck, A.J.R., Meijer, D.H. & Janssen, B.J.C. Structural insights into the contactin 1–neurofascin 155 adhesion complex. *Nature Communications* **13**, 6607 (2022).
13. Dingess, K.A., Hoek, M., van Rijswijk, D.M.H., Tamara, S., **den Boer, M.A.**, Veth, T., Damen, M.J.A., Barendregt, A., Romijn, M., Juncker, H.G., van Keulen, B.J., Vidarsson, G., van Goudoever, J.B., Bondt, A. & Heck, A.J.R. Identification of common and distinct origins of human serum and breastmilk IgA1 by mass spectrometry-based clonal profiling. *Cell. Mol. Immunol.*, 1-12 (2022).
14. Aguinagalde Salazar, L., **den Boer, M.A.**, Castenmiller, S.M., Zwarthoff, S.A., de Haas, C., Aerts, P.C., Beurskens, F.J., Schuurman, J., Heck, A.J.R., van Kessel, K. & Rooijakkers, S.H.M. Promoting Fc-Fc interactions between anti-capsular antibodies provides strong immune protection against Streptococcus pneumoniae. *eLife* **12**, e80669 (2023).
15. Bondt, A., Hoek, M., Dingess, K., Tamara, S., de Graaf, B., Peng, W., **den Boer, M.A.**, Damen, M., Zwart, C., Barendregt, A., van Rijswijck, D.M.H., Grobben, M., Tejjani, K., Rijswijk, J.v., Völlmy, F., Snijder, J., Fortini, F., Papi, A., Volta, C.A., Campo, G., Contoli, M., van Gils, M.J., Spadaro, S., Rizzo, P. & Heck, A.J.R. No patient is the same; lessons learned from antibody repertoire profiling in hospitalized severe COVID-19 patients. *medRxiv*, 2022.2012.2023.22283896 (2022).

Exploring Humoral Immune Responses by Mass Spectrometry
PhD Thesis, Utrecht University

Maurits den Boer, 2023, Utrecht, The Netherlands

ISBN: 978-94-6483-391-1