Title

Simultaneous polyclonal antibody sequencing and epitope mapping by cryo electron microscopy and mass spectrometry – a perspective.

Authors

Douwe Schulte[1], Marta Šiborová[1], Lukas Käll[2], Joost Snijder[1]*

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

[2] Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, Royal Institute of Technology – KTH, Solna, Sweden

* corresponding author: j.snijder@uu.nl

Abstract

Antibodies are a major component of adaptive immunity against invading pathogens. Here we explore possibilities for an analytical approach to characterize the antigen-specific antibody repertoire directly from the secreted proteins in convalescent serum. This approach aims to perform simultaneous antibody sequencing and epitope mapping using a combination of single particle cryo-electron microscopy (cryoEM) and bottom-up proteomics techniques based on mass spectrometry (LC-MS/MS). We evaluate the performance of the deep-learning tool ModelAngelo in determining *de novo* antibody sequences directly from reconstructed 3D volumes of antibody-antigen complexes. We demonstrate that while map quality is a critical bottleneck, it is possible to sequence antibody variable domains from cryoEM reconstructions with accuracies of up to 80-90%. While the rate of errors exceeds the typical levels of somatic hypermutation, we show that the ModelAngelo-derived sequences can be used to assign the used V-genes. This provides a functional guide to assemble *de novo* peptides from LC-MS/MS data more accurately and improves the tolerance to a background of polyclonal antibody sequences. Following this proof-of-principle, we discuss the feasibility and future directions of this approach to characterize antigen-specific antibody repertoires.

Introduction

Adaptive immunity to invading pathogens is mediated to an important degree by antibodies[1–4]. The available repertoire of antibodies is unique in individuals and constantly shifting in response to immunological pressure under which activation, selection, proliferation, maturation, and differentiation of the antibody-producing B-cells takes place[4–6]. Therefore, understanding the molecular mechanisms of antibody-mediated immunity requires knowledge of how a complex repertoire of polyclonal antibodies targets a diverse landscape of epitopes on their respective antigens.

The analytical challenge at hand is to resolve antigen-antibody interactions down to the pairwise contacts between specific amino acid residues in the epitope and paratope regions, respectively. This would enable reconstruction of the evolutionary pathways of somatic recombination and hypermutation that lead to high-affinity antigen binding. Conversely, this would also reveal how this selective immune pressure drives the evolutionary pathways of antigenic drift in targeted pathogens[7–10]. Knowledge of the precise antibody sequences, as well as near-atomic details of the epitope-paratope interaction, are thus prerequisites to understand the coevolution between replicating pathogens and antibody-mediated adaptive immunity in the host.

Current methods to determine the antigen-specific antibody repertoire rely on single (memory) B-cell sorting, followed by targeted sequencing of the coding mRNAs for the heavy and light chains[11,12]. This enables the production of recombinant monoclonal antibodies, whose epitopes may be mapped to near-atomic structural detail by X-ray crystallography and cryo electron microscopy (cryoEM). This approach has generated a wealth of information about antibody-antigen interactions, though it is biased by the limited pool of memory B-cells that it probes. Antibodies function as circulating glycoproteins in bodily fluid, secreted from plasma cells which are in turn derived from diverse pools of memory B cells located in various tissues, including bone marrow, spleen, lymph nodes, and only to a minor degree in blood[13–15]. Serological assays aimed to determine binding and neutralization titers specifically look at the secreted antibody in bodily fluid, and it remains an outstanding question how this ‘serum compartment’ of the antibody repertoire relates both qualitatively and quantitatively to the minor population of memory B-cells found in peripheral blood. This calls for new analytical approaches that can derive both antibody sequence and epitope information straight from the secreted antibody product.

Such approaches have been developed in recent years, based on mass spectrometry and electron microscopy. First, using a bottom-up proteomics approach, antibody-derived peptides can be sequenced *de novo* from fragmentation spectra and assembled into full heavy/light chain sequences[12,16]. Sequence accuracy is such that functional monoclonal antibodies can be reconstructed from the input data, and several reports have described successful sequencing efforts of human serum, milk, and urine-derived antibodies[17–33]. Second, both hydrogen-deuterium exchange mass spectrometry and electron microscopy have been used to resolve a complex landscape of epitopes targeted by polyclonal antibody mixtures[34–44]. Ward and colleagues have reported that with the latter approach, which they coined Electron Microscopy based Polyclonal Epitope Mapping (EMPEM), they obtained near-atomic resolution reconstructions by cryoEM to resolve side-chain densities in the epitope-paratope region[34]. This opens the possibility to derive antibody sequence information and integrate this into the structural modelling of the interaction. In essence, the reconstructions might reveal which antibodies from the complex polyclonal mixture bind which epitopes on the antigens. The improved resolutions of current cryoEM approaches thus allow for a type of visual proteomics in which protein identity (*i.e.* antibody sequence) may be directly inferred from the reconstructed 3D volumes[45–53]. While the pure sequencing accuracies from these approaches is obviously limited by resolution/map quality, many tools have been recently developed to infer protein identity in an automated fashion, including cryoID, DeepTracer-ID, *findmysequence*, and most recently ModelAngelo[54–57].

Here we explore the use of ModelAngelo to derive *de novo* antibody sequences from experimental cryoEM density maps of antibody-antigen complexes. We have previously developed the software tool Stitch, which sorts and assembles MS-derived peptide sequences into full heavy/light chain sequences across complex repertoires[58,59]. We adapted Stitch to perform the same task on ModelAngelo-derived *de novo* models and test the accuracy of the approach on a benchmark of 164 publicly available cryoEM maps of monoclonal antibody-antigen pairs. We demonstrate that map quality is a critical bottleneck, but that antibody sequences can be derived with up to 80-90% accuracy. We test the utility of these sequences for assigning the used V-genes, which together with reconstruction of CDRH3 may offer a useful guide to assemble more accurate MS-derived peptide sequences (see Figure 1). We show that such EM-derived templates indeed improve MS-based sequencing accuracy in the context of complex antibody mixtures and that publicly available EMPEM reconstructions are of sufficient quality to leverage this approach. This proof-

of-principle offers a promising perspective to integrate cryoEM and MS methods for a comprehensive characterization of the antibody repertoire on both sequence and epitope levels.

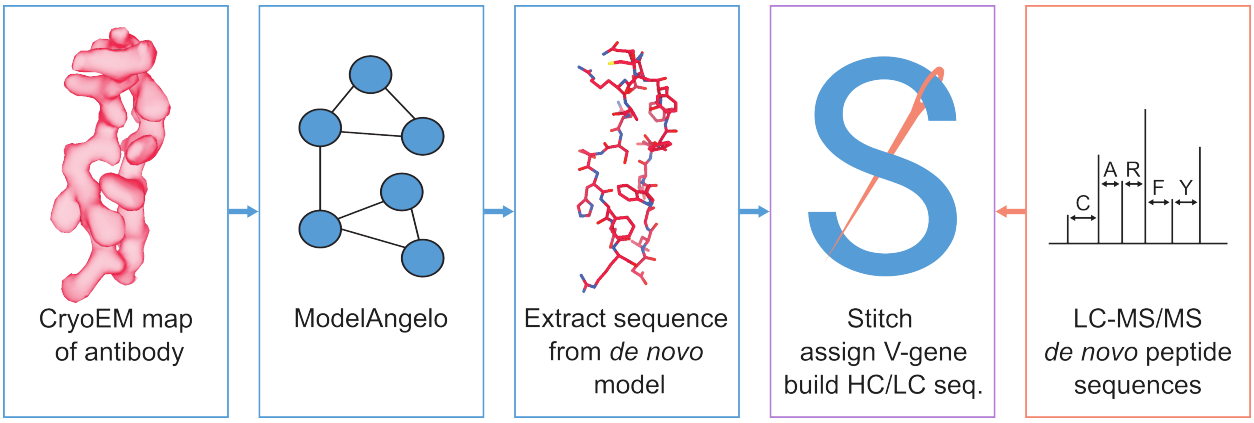

Figure 1. Schematic workflow to estimate de novo antibody sequences by the integration of cryoEM and LC-MS/MS data with Stitch.

Results

To assess the feasibility of deriving *de novo* antibody sequences from experimental cryoEM density maps, we assembled a benchmark dataset from the Electron Microscopy Data Bank (see Supplementary Table S1). To infer the antibody sequences, we chose the recently published deep-learning tool ModelAngelo, developed by Jamali, Scheres, and colleagues, as it is capable of inferring complete *de novo* models in cryoEM density maps without the need for user input sequences or main chain models[54–57]. We searched the EMDB for published maps containing antigen-binding fragments (Fabs) of any species, at nominal resolutions of ≤ 4 Å, released after the training data for ModelAngelo was obtained. These results were filtered for maps that included a deposited atomic model and which contained only a single unique monoclonal Fab (of which multiple copies may be bound in the reconstructed antigen complex). The final benchmark includes 164 maps, including Fabs from human, rabbit, mouse, and macaque species. The maps were used as input data for ModelAngelo without user-provided sequences, yielding completely *de novo* atomic models of both antigen and Fabs.

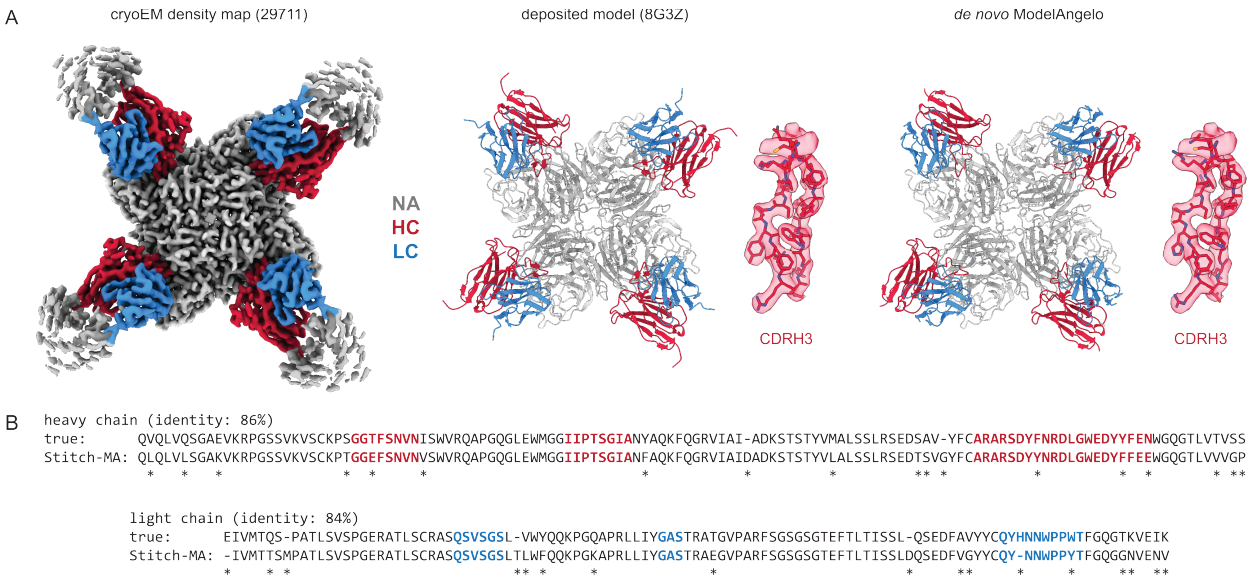

Figure 2. Determining de novo antibody sequences from cryoEM data with ModelAngelo. A) Exemplary map (with top10% alignment scores of 1076/1174 for HC/LC) from the benchmark dataset, representing an Influenza B virus neuraminidase (NA) in complex with four copies of a neutralizing Fab, at global FSC resolution of 2.3 Å. Shown are the deposited map, model, and the de novo model generated by ModelAngelo, along with a detailed view of CDRH3. B) Consensus sequences for heavy and light chains as generated by Stitch compared to the true sequences. Sequencing errors are indicated by an asterisk (*).

The output models from ModelAngelo are typically fragmented to varying degrees depending on the local quality of the map. In addition, maps may contain multiple copies of the same unique Fab molecule. We therefore aimed to consolidate all fragments for the built Fabs into a single consensus sequence for the antibody variable domains of the heavy and light chains. We have previously developed the software Stitch, which performs assembly of LC-MS/MS derived *de novo* peptide reads into the correct framework of the heavy and light chains by alignment to germline-template sequences from the ImMunoGeneTics (IMGT) database[60–62]. We adapted Stitch to use models from ModelAngelo (or any mmCIF file) as input data, extract the amino acid sequences, and perform the same template-based assembly with the resulting reads. Of the 164 input maps, 141 and 144 yielded a non-zero alignment score for the heavy and light chains in the Stitch result, respectively. These add up to a total of 152 maps which were analyzed further to evaluate the sequence accuracy of this approach (including 134 maps with a non-zero alignment score for *both* heavy and light chain).

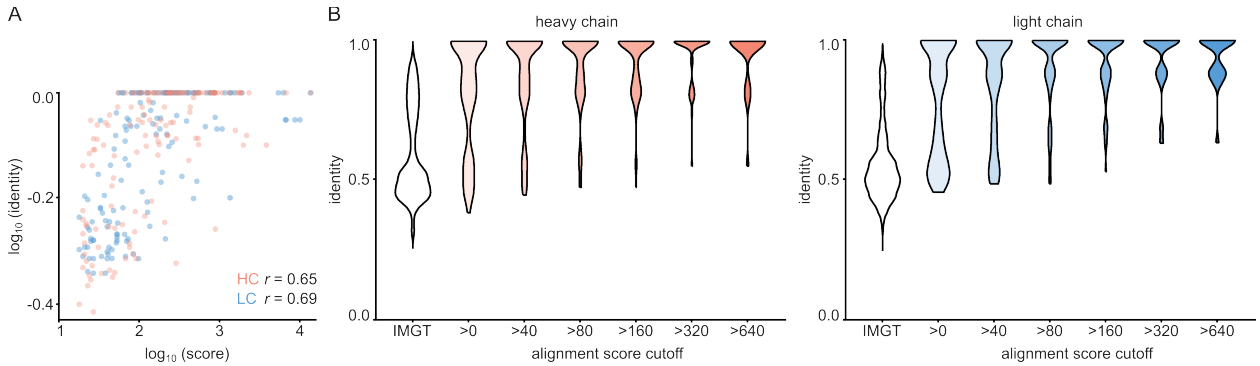

Figure 3. V-gene assignment from ModelAngelo data. A) Correlation between Stitch alignment score and sequence identity between the top-scoring V-gene of the ModelAngelo vs PDB sequence of the heavy and light chain variable domains, as indicated with the non-parametric Spearman correlation coefficient. B) Distribution of V-gene sequence identity for progressive alignment score cutoffs, compared to the pairwise V-gene sequence identity in the IMGT repertoire.

Results for one of the top-scoring entries in the benchmark, an Influenza B virus neuraminidase tetramer in complex with four identical copies of a neutralizing Fab[63], is shown in Figure 2. The consensus sequences generated in Stitch have complete coverage of both heavy and light chain variable domains, including both CDRL3 and CDRH3. The *de novo* determined sequence is 84% and 86% identical to the true heavy and light chain sequences, respectively. The 15% rate of sequencing errors exceeds the typical levels of somatic hypermutation observed in mature antibody sequences, which are on the order of 1-10%. While the derived sequence should therefore not be taken at face value to reconstruct recombinant monoclonal antibodies, we reasoned that the accuracy is nevertheless likely sufficient to correctly infer the corresponding germline V-genes of the mature antibodies.

For each of the 152 maps, we compared the pairwise sequence identity between the top scoring V-genes from the ModelAngelo input in Stitch with the top scoring V-genes from the true sequences (as deposited in the corresponding PDB entry). For reference, we also calculated the pairwise sequence identities of all available V-gene templates per species, reflecting what a completely random draw from the available V-gene repertoire would look like. As shown in Figure 3, the top scoring V-genes from the ModelAngelo sequences have significantly higher pairwise identities than a random draw from the V-gene repertoire in IMGT for both the heavy and light chains ($p < 0.0001$ in unpaired, two-tailed Kolmogorov Smirnov tests). Furthermore, the identity of the inferred V-gene with the true sequence scales with the alignment scores in Stitch, making it a valuable metric for the quality of the V-gene inference. The mean/median V-gene identity in a

random draw from the IMGT repertoire is 0.59/0.52 and 0.56/0.53 for the heavy and light chains, respectively. With the ModelAngelo-derived sequences this improves to 0.78/0.87 and 0.75/0.85 for the heavy and light chains in the full dataset. This gradually improves with higher alignment scores to 0.89/1.00 and 0.88/1.00 for heavy and light chains, starting at a cutoff of 80 (representing 92/141 and 80/134 maps for heavy and light chain, respectively). The complete complementarity determining regions CDRH3 and CDRL3 were covered in 66 and 68 maps for the heavy and light chain, respectively (see Supplementary Figure S1). The length of the complete antigen binding loops was estimated with an average error of 0.5 ± 3.3 or 1.7 ± 6.0 residues for heavy and light chain, with average sequence identities of 0.63 and 0.41. While CDRH3 is the more challenging region in MS-based approaches to antibody sequencing, we believe that the moderately better length and sequence accuracy of CDRH3 compared to CDRL3 in ModelAngelo output reflects the CDRH3's notoriously tight involvement in antigen binding, hence a greater relative stability in the antibody-antigen complex, resulting in better order in the reconstructed EM density maps. We found the global FSC resolution of the input map to be a poor predictor of both the Stitch alignment score and the inferred V-gene identity, likely because it is dominated by the bulk of the antigen and not representative of the local resolution in the epitope-paratope region (see Supplementary Figure S2). These results demonstrate that candidate V-genes for the antibodies resolved in cryoEM densities can be accurately narrowed down using ModelAngelo and Stitch, and that a limited subset of maps contains accurate information on CDR3 sequence and length.

In the context of polyclonal antibody mixtures, this analysis suggests that cryoEM densities of antigen-antibody complexes from EMPEM experiments can be leveraged to guide sequence assembly from complementary proteomics-based profiling of the same sample (see Figure 1). In such an experiment, reconstructed cryoEM densities would be used as input data for ModelAngelo, from which the sequences are extracted and run through Stitch to select the top-scoring V-gene and construct a placeholder sequence for CDR3 of both the heavy and light chain. These reconstructed variable domains may then act as templates to guide the assembly of *de novo* peptides from LC-MS/MS data to improve the accuracy of the candidate sequence.

As a proof-of-principle, we tested this on the monoclonal antibody CR3022, for which both a cryoEM reconstruction and LC-MS/MS data are publicly available. This antibody was isolated from a convalescent survivor of a SARS-CoV infection and targets a cryptic, but conserved epitope at the base of the Spike receptor-binding domain, with cross-neutralization to SARS-CoV-

2[64,65]. The antibody consists of an IGHV5-51 heavy chain, paired with an IGKV4-1 light chain. When complexed with full-length SARS-CoV-2 Spike, its Fab induces an odd rearrangement of the Spike protomers to yield an antiparallel dimer of S1 subunits in the cryoEM reconstructions[66,67]. When using this map as input for ModelAngelo and subsequently Stitch, the IGHV5-51 and IGKV4-1 germline sequences are correctly identified based on their alignment scores. Moreover, complete sequences for CDR3 of the heavy and light chains are built. When using these reconstructed variable domains as templates to guide assembly of the *de novo* peptide reads from the LC-MS/MS data published by Person and colleagues[68], the final consensus sequences are 96% and 99% identical to the true heavy and light chain respectively. Of note, the only three remaining errors in the six CDR sequences are I/L assignments, which have identical masses and are notoriously challenging for MS-based sequencing.

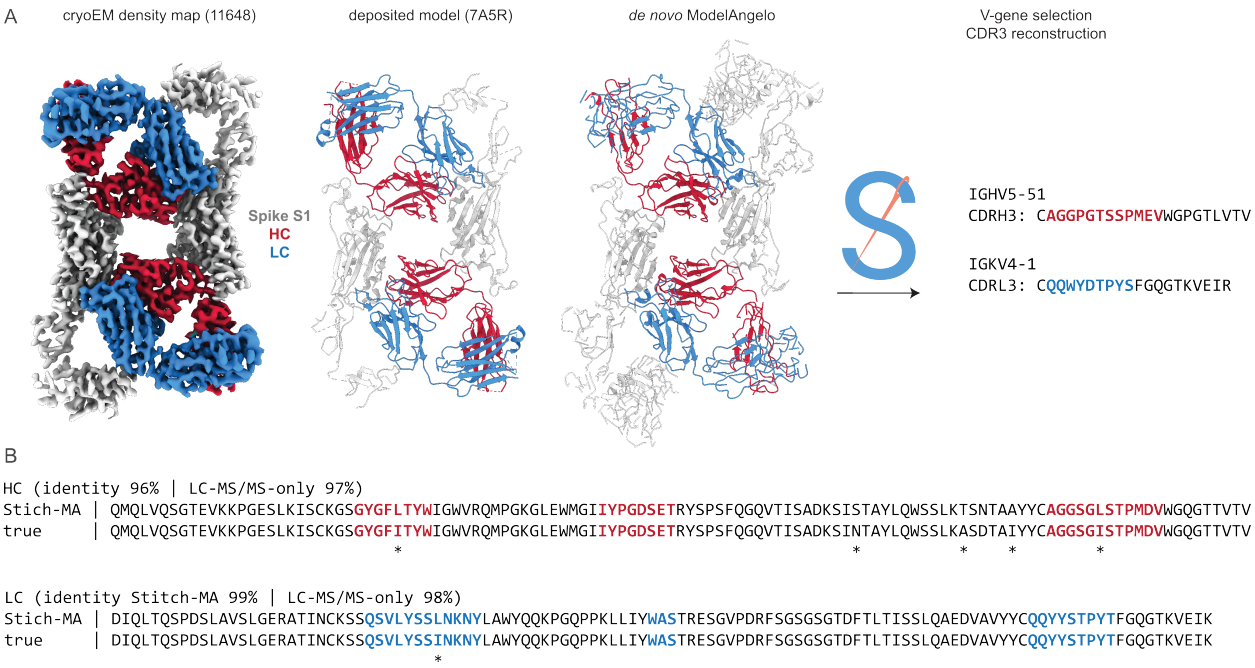

Figure 4. Sequencing CR3022 with integrated cryoEM and LC-MS/MS data. A) Shown are the deposited cryoEM map (global FSC resolution 4.1 Å), model, and de novo ModelAngelo output for the CR3022 Fab in complex with the SARS-CoV-2 Spike S1 subunit. The sequences were extracted from the de novo model and used as input for Stitch, resulting in the identification of the indicated V-genes and CDR3 sequences. These variable domains were used as templates in Stitch to assemble the LC-MS/MS derived de novo peptides. B) Consensus sequences for CR3022 from the integrated cryoEM-MS data in Stitch compared to the true heavy and light chain sequences. Sequencing errors are indicated with an asterisk (*).

For this CR3022 dataset, derived from the monoclonal antibody, the LC-MS/MS data alone assembled with Stitch against the full range of V-genes already yields a similar accuracy of 97%

and 98% for the heavy and light chain, respectively. For the case of an EMPEM experiment, the challenge would rather be to correctly assemble the CR3022 sequence against a background of unrelated antibody sequences in a complex mixture. We therefore tested the utility of the ModelAngelo-derived templates for sequence assembly by also mixing input reads from LC-MS/MS data of unrelated antibodies (Figure 4). First, we mixed in a background of whole IgG from a hospitalized COVID-19 patient[58], representing a diffuse polyclonal background in approximately a 1:1 ratio to the target input reads. Second, we mixed in a background of five additional unrelated anti-Influenza-HA monoclonal antibodies from the same study of Person and colleagues[68], amounting to a 5:1 ratio of background to target input reads. In our previous work on sequencing serum-derived antibodies by bottom-up proteomics with Stitch, we found that assembly of the consensus sequences becomes much more tolerant to background data if, beyond the top scoring V-gene, the remaining unrelated V-genes are also included as decoys for the final template matching step[58]. The use of these decoy sequences is also included in the comparison here. Furthermore, we also included the use of the true CR3022 sequences as templates, serving as a best-case scenario, positive control. The analysis confirmed that even without ModelAngelo-derived templates, the sequence assembly with Stitch is already tolerant to the diffuse polyclonal background from the whole IgG fraction, yielding a similar accuracy as in the absence of these background peptides. By contrast, the sequence accuracy plummets to below 0.6 when the background consists of the five additional monoclonal antibodies in a 5:1 ratio to the target input reads. The accuracy is recovered to >0.95 when using either the true CR3022 sequences or the ModelAngelo-derived templates. There is also a gain in accuracy by using decoy templates with the LC-MS/MS data alone, though smaller compared to the use of ModelAngelo-derived templates. These results demonstrate that ModelAngelo-derived templates are useful for sequence assembly against complex polyclonal backgrounds.

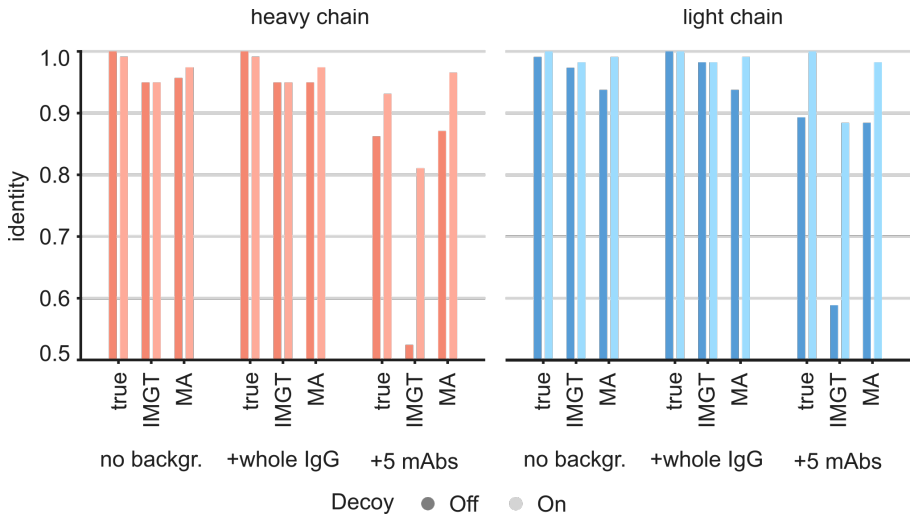

Figure 5. Targeted sequencing of CR3022 against a complex background of other antibodies. Plotted are the de novo consensus sequence identities derived from the LC-MS/MS data using either the true sequences, the full IMGT repertoire, or the ModelAngelo-derived variable domains as templates. We compare the output from the CR3022 dataset alone ('No backgr.') with the output after adding either a diffuse polyclonal IgG background from a COVID-19 patient ('+whole IgG') or full datasets from five additional anti-Influenza-HA monoclonals ('+5 mAbs'). Use of decoy sequences as indicated by dark/light colors.

We have demonstrated that cryoEM reconstructions of monoclonal antigen-antibody complexes may contain sufficient information to accurately narrow down candidate V-genes and that this can be integrated with proteomics data to improve the accuracy of candidate sequences. We also evaluated whether EMPEM data is indeed of sufficient quality to infer V-genes from automated *de novo* modeling in the maps. We downloaded all 46 published EMPEM maps from EMDB, of which 23 were of sufficient quality to give a non-zero alignment score when using the ModelAngelo results in Stitch (see Supplementary Table S2). Analysis of the benchmark of monoclonal antigen-antibody complexes showed that the quality of the V-gene inference scales with the alignment scores. In this set of EMPEM reconstructions, the alignment scores range from ca. 20 to 300. Of the 23 maps, 8 have alignment scores above 50 for both heavy and light chain, at which point we estimate the mean/median V-gene identity to be approximately 0.85/0.90. This analysis shows that experimental EMPEM studies may yield sufficiently detailed reconstructions of the antigen-antibody complexes to narrow down the candidate V-genes of the resolved Fabs accurately.

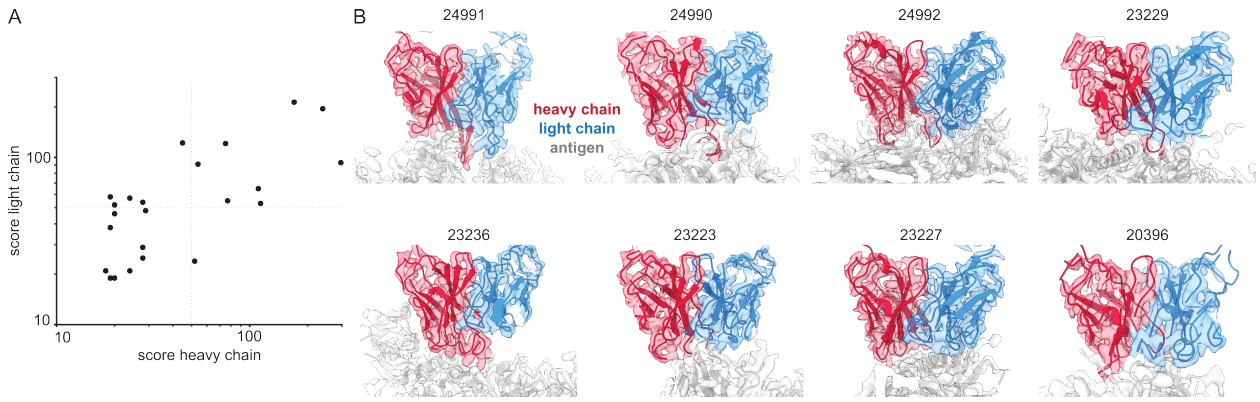

Figure 6. Inferring V-genes from published EMPEM data. A) Plotted are all non-zero alignment scores in Stitch from published EMPEM maps. B) Views of the variable domains of EMPEM maps with alignment scores >50 for both heavy and light chains. The EMDB identifiers are indicated at each panel.

Discussion

The development of EMPEM and MS-based polyclonal antibody sequencing now make it possible to profile the antigen-specific antibody repertoire straight from the secreted pool of antibodies in bodily fluids. This approach bridges the gap between established single B-cell sequencing approaches and serological assays to probe binding and neutralization titers. While sample complexity remains an important bottleneck, and questions remain about the dynamic range of the true serum antibody repertoire and the depth of coverage from these novel experimental approaches, several studies have recently reached the important milestone of reconstructing functional antibodies from direct measurements of the secreted serum components[17,19,20,22,23,25,28,30,34,69]. The present work demonstrates that epitope and sequence information can be integrated using ModelAngelo and Stitch. This approach holds promise to better understand the serum compartment of the antibody repertoire. Use of the cryoEM data in these workflows complements the MS data beyond epitope mapping in several significant ways. First, we demonstrated that narrowing down the candidate V-genes improves the tolerance of LC-MS/MS-derived peptide sequence assembly to background in a complex antibody mixture. Second, heavy-light chain pairing is a problematic blind spot to proteomics sequencing, as the antibodies are denatured and digested as part of the sample workup. In contrast, this pairing is trivial in cryoEM data as the chains are in direct contact in the resolved Fabs in the map. Finally, reconstruction of CDRH3 is especially challenging with proteomics data alone, as it spans the junction between the recombined V-, D-, and J-segments, of which the D-segment is short and hypervariable to a point that germline sequences do not provide a functional template for

290 sequence assembly. While CDRH3 coverage is also limited in ModelAngelo data, it can be built
291 in many cases and a correct estimate of CDRH3 length is already useful to guide assembly of the
292 *de novo* peptide reads.

293 Here we have extracted the flat sequences from ModelAngelo output to use as input for the
294 template matching step in Stitch, using a modified Smith-Waterman Alignment. We believe the
295 template matching step in Stitch could be further improved in several ways. First, ModelAngelo
296 has a built-in database search based on *HMMer* using profile Hidden Markov models (HMM)
297 encoding the amino acid probabilities across the alphabet at each position[70]. While the
298 ModelAngelo search currently does not consolidate the fragmented output models into a single
299 search to build consensus sequences, we anticipate that implementing a similar HMM profile
300 search in Stitch may further improve the V-gene inference. Indeed, a recent report by Ward and
301 colleagues demonstrates that ModelAngelo and its *HMMer* search could be implemented to
302 correlate high-resolution EMPEM data with B-cell receptor mRNA sequencing data to speed up
303 and improve success rates of antibody discovery compared to previously reported data analysis
304 pipelines[69]. In addition, from the benchmark dataset analyzed here we might learn what
305 sequencing mistakes are common in ModelAngelo data. This could be used to adjust the
306 conventional Smith-Waterman Alignment in Stitch accordingly, analogous to recent
307 improvements in the alignment algorithm we implemented for MS data[59]. Finally, the template
308 matching step in Stitch is now solely based on the sequence of the ModelAngelo models, but we
309 may expand this to a structure-based alignment to better place the error-prone sequence reads
310 in the correct framework of the Ig-domains.

311 Next steps in our efforts are to bring together the EMPEM work and MS-based polyclonal antibody
312 sequencing on antigen-Fab complexes purified from patient serum. The goal is to reconstruct
313 functional monoclonal antibodies from these analyses as ultimate proof for the accuracy of the
314 derived antibody sequences and then start working on the throughput and depth of coverage in
315 the repertoire. We believe that both EMPEM and MS-based polyclonal antibody sequencing are
316 still limited to the top 1-10 antibodies in the polyclonal mixture. The EMPEM approach is biased
317 towards bigger and well-ordered target antigens, which calls for additional complementary
318 approaches like HDX-MS for a comprehensive polyclonal epitope mapping exercise. Bringing
319 together these perspectives in an integrated structural biology approach promises new insights
320 into the serum compartment of the antibody repertoire to better understand the coevolutionary
321 processes of antibody maturation and antigenic drift.

Materials and Methods

Benchmark of monoclonal antibody-antigen pairs and EMPEM maps from EMDB

We searched the EMDB for published maps containing antigen-binding fragments (Fabs) of any species, at nominal resolutions of ≤ 4 Ångström, released after the training data for ModelAngelo was obtained (April 1st 2022). The search was performed on February 11th 2023 at https://www.ebi.ac.uk/emdb/ using the search term "antibody fab resolution:[* TO 4] AND release_date:[2022-03-31T00:00:00Z TO 2023-11-02T00:00:00Z] sample_name:*fab* fitted_pdbs:[* TO *]". These results were filtered for maps that included a deposited atomic model and which contained only a single unique monoclonal Fab (of which multiple copies may be bound in the reconstructed antigen complex). When multiple redundant maps from the same study were included, we selected the single representative map of the highest quality, based on manual inspection. Global FSC resolution was not a good indicator as the maps were often dominated by the bound antigen, which may be better resolved than the bound antibody. Typically, the selected map was the focused/local refinement around the epitope-paratope region, despite its lower nominal resolution. A full overview of the selected maps is provided in Supplementary Table S1. The EMPEM maps were compiled based on a literature survey, complemented with a search of the EMDB using the term "polyclonal". A full overview of the selected EMPEM maps is provided in Supplementary Table S2.

Changes made to Stitch to take CIF input

Stitch was extended to allow mmCIF files as input. From these files all chains were extracted as separate amino acid sequences to align in Stitch. ModelAngelo outputs a confidence score per residue in the B-factor column of the mmCIF input, which was used as a local confidence for the sequence. First, each polypeptide gets assigned an Average Local Confidence (ALC) score based on the average across all residues, which can be used as an input filter on the data (along with polypeptide length). Second, the local confidence is used as weight in determining the consensus sequence of overlapping polypeptides, following assembly in Stitch.

Analysis of monoclonal antigen-antibody and EMPEM benchmarks

Software used in this project was curated by SBGrid[71]. The deposited model mmCIF and EM maps for the full benchmark were downloaded and ran with ModelAngelo (version 1.01) in 'build_no_seq' mode. For each entry in the benchmark the deposited mmCIF was run with Stitch (version 1.5.0-rc.1+6d3b540) using CutoffALC 80, minimum length 5 and TemplateMatching

CutoffScore 8. The same was done for each mmCIF file produced by ModelAngelo but with an additional segment containing the antigen and Ig constant domain template sequences. From these runs the consensus sequence and highest scoring germline for IGHV and IGLV (lambda + kappa) were retrieved. For each consensus sequence the CDR3 was determined if the flanking cysteine on the V gene and the tryptophan or phenylalanine on the J gene were present. As these conserved residues were not all positioned correctly in the IMGT database the data was manually fixed based on the same rules. The data from the deposited and produced Stitch runs was compared to produce the identity between the consensus sequences, distance between the inferred germlines, and identity between the CDR3s. The script used for this analysis is deposited as Supplementary Data. The results generated by this analysis is included in Supplementary Table S1. The EMPEM benchmark was downloaded, ran through ModelAngelo and subsequently Stitch with identical parameters as above. In contrast to the benchmark detailed before, the ground-truth sequences of these antibodies is not known.

Analysis of CR3022 data

The EM data for CR3022 was downloaded from EMDB (EMD-11648) and run with ModelAngelo (version 1.01) in 'build_no_seq' mode. The raw data for monoclonal antibodies CR3022, 107, 1028, 2771, 3576, and 3634 from PRIDE PXD030094 was downloaded and analyzed with PEAKS 10+. These listed monoclonal antibodies were chosen because these are the five most distant sequences from CR3022 and therefore present the biggest challenge to sequence assembly in Stitch (version 1.5.0-rc.1+6d3b540). The PEAKS analysis for the COVID-19 data from PRIDE PXD031941 was downloaded. Three sets of input were prepared: 'no background' consisting of only the CR3022 data, '+whole IgG' consisting of the CR3022 and the COVID-19 data, and '+5 mAbs' which consists of all mAbs from the CR3022 study. Three Stitch configurations where prepared: 'True' using the known CR3022 sequence as template as retrieved from PDB 7A5R, 'IMGT' using the conventional configuration of Stitch with all IMGT germlines as templates, and 'MA' using the closest V gene germline to the ModelAngelo consensus sequence of CR3022 together with the CDR3 sequence present. Each of these configurations were run with Stitch Recombine Decoy off and on, with this on for 'True' and 'MA' the full IMGT germline database was added and for 'IMGT' the Stitch parameter Decoy in Recombine was set allowing any unused germline from the Template Matching step to matched in the Recombine step. For all Stitch runs the CutoffALC was 90 and the TemplateMatching CutoffScore 10. The resulting consensus sequences from these 18 Stitch runs were then compared with the known CR3022 sequence to

determine the identity. The full script used for this analysis is included in the deposited Supplementary Data.

Data and code availability

Stitch is available at https://github.com/snijderlab/stitch. The CR3022 and COVID-19 whole IgG LC-MS/MS data were taken from PRIDE Archive (http://www.ebi.ac.uk/pride/archive/) via the PRIDE partner repository with the data set identifiers PXD030094 and PXD031941, respectively. All ModelAngelo and Stitch results, including a script to reproduce the full analysis, are made available on Zenodo under 10.5281/zenodo.12207014.

Supplementary information

Table S1: Overview and results of maps for benchmark.

Table S2: Overview and results of maps for EMPEM benchmark.

Figure S1: Overview of CDR3 results from benchmark.

Figure S2: Correlation of global FSC resolution with results from benchmark.

Acknowledgements

This research was funded by the Dutch Research Council NWO Gravitation 2013 BOO, Institute for Chemical Immunology (ICI; 024.002.009), and the European Research Council Executive Agency HORIZON ERC-2022-STG (FLAVIR; 101077640).

References

(1) Bonilla, F. A.; Oettgen, H. C. Adaptive Immunity. *J. Allergy Clin. Immunol.* **2010**, *125* (2), S33–S40. https://doi.org/10.1016/j.jaci.2009.09.017.
(2) Rees, A. R. Understanding the Human Antibody Repertoire. *mAbs* **2020**, *12* (1), 1729683. https://doi.org/10.1080/19420862.2020.1729683.
(3) Burton, D. R. Antiviral Neutralizing Antibodies: From in Vitro to in Vivo Activity. *Nat. Rev. Immunol.* **2023**, *23* (11), 720–734. https://doi.org/10.1038/s41577-023-00858-w.
(4) Lam, J. H.; Smith, F. L.; Baumgarth, N. B Cell Activation and Response Regulation During Viral Infections. *Viral Immunol.* **2020**, *33* (4), 294–306. https://doi.org/10.1089/vim.2019.0207.
(5) Tellier, J.; Nutt, S. L. Plasma Cells: The Programming of an Antibody-secreting Machine. *Eur. J. Immunol.* **2019**, *49* (1), 30–37. https://doi.org/10.1002/eji.201847517.
(6) Pieper, K.; Grimbacher, B.; Eibel, H. B-Cell Biology and Development. *J. Allergy Clin. Immunol.* **2013**, *131* (4), 959–971. https://doi.org/10.1016/j.jaci.2013.01.046.
(7) Marks, C.; Deane, C. M. How Repertoire Data Are Changing Antibody Science. *J. Biol. Chem.* **2020**, *295* (29), 9823–9837. https://doi.org/10.1074/jbc.REV120.010181.
(8) Han, A. X.; De Jong, S. P. J.; Russell, C. A. Co-Evolution of Immunity and Seasonal Influenza Viruses. *Nat. Rev. Microbiol.* **2023**, *21* (12), 805–817. https://doi.org/10.1038/s41579-023-00945-8.

(9) Vajda, S.; Porter, K. A.; Kozakov, D. Progress toward Improved Understanding of Antibody Maturation. *Curr. Opin. Struct. Biol.* **2021**, *67*, 226–231. https://doi.org/10.1016/j.sbi.2020.11.008.
(10) White, H. N. B-Cell Memory Responses to Variant Viral Antigens. *Viruses* **2021**, *13* (4), 565. https://doi.org/10.3390/v13040565.
(11) Georgiou, G.; Ippolito, G. C.; Beausang, J.; Busse, C. E.; Wardemann, H.; Quake, S. R. The Promise and Challenge of High-Throughput Sequencing of the Antibody Repertoire. *Nat. Biotechnol.* **2014**, *32* (2), 158–168. https://doi.org/10.1038/nbt.2782.
(12) 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.* **2015**, *24*, 112–120. https://doi.org/10.1016/j.cbpa.2014.11.007.
(13) Inoue, T.; Kurosaki, T. Memory B Cells. *Nat. Rev. Immunol.* **2024**, *24* (1), 5–17. https://doi.org/10.1038/s41577-023-00897-3.
(14) Meng, W.; Zhang, B.; Schwartz, G. W.; Rosenfeld, A. M.; Ren, D.; Thome, J. J. C.; Carpenter, D. J.; Matsuoka, N.; Lerner, H.; Friedman, A. L.; Granot, T.; Farber, D. L.; Shlomchik, M. J.; Hershberg, U.; Luning Prak, E. T. An Atlas of B-Cell Clonal Distribution in the Human Body. *Nat. Biotechnol.* **2017**, *35* (9), 879–884. https://doi.org/10.1038/nbt.3942.
(15) Akkaya, M.; Kwak, K.; Pierce, S. K. B Cell Memory: Building Two Walls of Protection against Pathogens. *Nat. Rev. Immunol.* **2020**, *20* (4), 229–238. https://doi.org/10.1038/s41577-019-0244-2.
(16) 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* **2022**, *14* (1), 2079449. https://doi.org/10.1080/19420862.2022.2079449.
(17) Fridy, P. C.; Li, Y.; Keegan, S.; Thompson, M. K.; Nudelman, I.; Scheid, J. F.; Oeffinger, M.; Nussenzweig, M. C.; Fenyö, D.; Chait, B. T.; Rout, M. P. A Robust Pipeline for Rapid Production of Versatile Nanobody Repertoires. *Nat. Methods* **2014**, *11* (12), 1253–1260. https://doi.org/10.1038/nmeth.3170.
(18) Peng, W.; Giesbers, K. C. A. P.; Šiborová, M.; Beugelink, J. W.; Pronker, M. F.; Schulte, D.; Hilkens, J.; Janssen, B. J. C.; Strijbis, K.; Snijder, J. Reverse Engineering the Anti-MUC1 Hybridoma Antibody 139H2 by Mass Spectrometry-Based de Novo Sequencing. *bioRxiv* **2023**, 2023.07.05.547778. https://doi.org/10.26508/lsa.202302366.
(19) 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.; Schulte, D.; Grobben, M.; Tejjani, K.; van Rijswijk, J.; 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. Into the Dark Serum Proteome: Personalized Features of IgG1 and IgA1 Repertoires in Severe COVID-19 Patients. *Mol. Cell. Proteomics* **2024**, *23* (1), 100690. https://doi.org/10.1016/j.mcpro.2023.100690.
(20) 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**, *22* (9), 3022–3028. https://doi.org/10.1021/acs.jproteome.3c00330.
(21) Tran, N. H.; Rahman, M. Z.; He, L.; Xin, L.; Shan, B.; Li, M. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **2016**, *6* (1), 31730. https://doi.org/10.1038/srep31730.
(22) 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 Syst.* **2021**, *12* (12), 1131-1143.e5. https://doi.org/10.1016/j.cels.2021.08.008.

(23) Sousa, E.; Olland, S.; Shih, H. H.; Marquette, K.; Martone, R.; Lu, Z.; Paulsen, J.; Gill, D.; He, T. Primary Sequence Determination of a Monoclonal Antibody against α-Synuclein Using a Novel Mass Spectrometry-Based Approach. *Int. J. Mass Spectrom.* **2012**, *312*, 61–69. https://doi.org/10.1016/j.ijms.2011.05.005.

(24) Sen, K. I.; Tang, W. H.; Nayak, S.; Kil, Y. J.; Bern, M.; Ozoglu, B.; Ueberheide, B.; Davis, D.; Becker, C. Automated Antibody De Novo Sequencing and Its Utility in Biopharmaceutical Discovery. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (5), 803–810. https://doi.org/10.1007/s13361-016-1580-0.

(25) Rickert, K. W.; Grinberg, L.; Woods, R. M.; Wilson, S.; Bowen, M. A.; Baca, M. Combining Phage Display with de Novo Protein Sequencing for Reverse Engineering of Monoclonal Antibodies. *mAbs* **2016**, *8* (3), 501–512. https://doi.org/10.1080/19420862.2016.1145865.

(26) Savidor, A.; Barzilay, R.; Elinger, D.; Yarden, Y.; Lindzen, M.; Gabashvili, A.; Adiv Tal, O.; Levin, Y. Database-Independent Protein Sequencing (DiPS) Enables Full-Length de Novo Protein and Antibody Sequence Determination. *Mol. Cell. Proteomics* **2017**, *16* (6), 1151–1161. https://doi.org/10.1074/mcp.O116.065417.

(27) 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.* **2021**, *20* (7), 3559–3566. https://doi.org/10.1021/acs.jproteome.1c00169.

(28) Guthals, A.; Gan, Y.; Murray, L.; Chen, Y.; Stinson, J.; Nakamura, G.; Lill, J. R.; Sandoval, W.; Bandeira, N. De Novo MS/MS Sequencing of Native Human Antibodies. *J. Proteome Res.* **2017**, *16* (1), 45–54. https://doi.org/10.1021/acs.jproteome.6b00608.

(29) Cheung, W. C.; Beausoleil, S. A.; Zhang, X.; Sato, S.; Schieferl, S. M.; Wieler, J. S.; Beaudet, J. G.; Ramenani, R. K.; Popova, L.; Comb, M. J.; Rush, J.; Polakiewicz, R. D. A Proteomics Approach for the Identification and Cloning of Monoclonal Antibodies from Serum. *Nat. Biotechnol.* **2012**, *30* (5), 447–452. https://doi.org/10.1038/nbt.2167.

(30) Castellana, N. E.; McCutcheon, K.; Pham, V. C.; Harden, K.; Nguyen, A.; Young, J.; Adams, C.; Schroeder, K.; Arnott, D.; Bafna, V.; Grogan, J. L.; Lill, J. R. Resurrection of a Clinical Antibody: Template Proteogenomic de Novo Proteomic Sequencing and Reverse Engineering of an Anti-lymphotoxin-α Antibody. *PROTEOMICS* **2011**, *11* (3), 395–405. https://doi.org/10.1002/pmic.201000487.

(31) Bandeira, N.; Pham, V.; Pevzner, P.; Arnott, D.; Lill, J. R. Automated de Novo Protein Sequencing of Monoclonal Antibodies. *Nat. Biotechnol.* **2008**, *26* (12), 1336–1338. https://doi.org/10.1038/nbt1208-1336.

(32) Dupré, M.; Duchateau, M.; Sternke-Hoffmann, R.; Boquoi, A.; Malosse, C.; Fenk, R.; Haas, R.; Buell, A. K.; Rey, M.; Chamot-Rooke, J. *De Novo* Sequencing of Antibody Light Chain Proteoforms from Patients with Multiple Myeloma. *Anal. Chem.* **2021**, *93* (30), 10627–10634. https://doi.org/10.1021/acs.analchem.1c01955.

(33) Peng, W.; Šiborová, M.; Wu, X.; Du, W.; Schulte, D.; Pronker, M. F.; De Haan, C. A. M.; Snijder, J. Structural Basis for Postfusion-Specific Binding to Respiratory Syncytial Virus F Protein by the Canonical Antigenic Site I Antibody 131-2a. Immunology January 4, 2025. https://doi.org/10.1101/2025.01.04.631317.

(34) Antanasijevic, A.; Bowman, C. A.; Kirchdoerfer, R. N.; Cottrell, C. A.; Ozorowski, G.; Upadhyay, A. A.; Cirelli, K. M.; Carnathan, D. G.; Enemuo, C. A.; Sewall, L. M.; Nogal, B.; Zhao, F.; Groschel, B.; Schief, W. R.; Sok, D.; Silvestri, G.; Crotty, S.; Bosinger, S. E.; Ward, A. B. From Structure to Sequence: Antibody Discovery Using cryoEM. *Sci. Adv.* **2022**, *8* (3), eabk2039. https://doi.org/10.1126/sciadv.abk2039.

(35) Antanasijevic, A.; Schulze, A. J.; Reddy, V. S.; Ward, A. B. High-Resolution Structural Analysis of Enterovirus-Reactive Polyclonal Antibodies in Complex with Whole Virions. *PNAS Nexus* **2022**, *1* (5), pgac253. https://doi.org/10.1093/pnasnexus/pgac253.

(36) Bangaru, S.; Antanasijevic, A.; Kose, N.; Sewall, L. M.; Jackson, A. M.; Suryadevara, N.; Zhan, X.; Torres, J. L.; Copps, J.; De La Peña, A. T.; Crowe, J. E.; Ward, A. B. Structural Mapping of Antibody Landscapes to Human Betacoronavirus Spike Proteins. *Sci. Adv.* **2022**, *8* (18), eabn2911. https://doi.org/10.1126/sciadv.abn2911.

(37) Boyoglu-Barnum, S.; Ellis, D.; Gillespie, R. A.; Hutchinson, G. B.; Park, Y.-J.; Moin, S. M.; Acton, O. J.; Ravichandran, R.; Murphy, M.; Pettie, D.; Matheson, N.; Carter, L.; Creanga, A.; Watson, M. J.; Kephart, S.; Ataca, S.; Vaile, J. R.; Ueda, G.; Crank, M. C.; Stewart, L.; Lee, K. K.; Guttman, M.; Baker, D.; Mascola, J. R.; Veesler, D.; Graham, B. S.; King, N. P.; Kanekiyo, M. Quadrivalent Influenza Nanoparticle Vaccines Induce Broad Protection. *Nature* **2021**, *592* (7855), 623–628. https://doi.org/10.1038/s41586-021-03365-x.

(38) Dingens, A. S.; Pratap, P.; Malone, K.; Hilton, S. K.; Ketas, T.; Cottrell, C. A.; Overbaugh, J.; Moore, J. P.; Klasse, P.; Ward, A. B.; Bloom, J. D. High-Resolution Mapping of the Neutralizing and Binding Specificities of Polyclonal Sera Post-HIV Env Trimer Vaccination. *eLife* **2021**, *10*, e64281. https://doi.org/10.7554/eLife.64281.

(39) Grauslund, L. R.; Ständer, S.; Veggi, D.; Andreano, E.; Rand, K. D.; Norais, N. Epitope Mapping of Human Polyclonal Antibodies to the fHbp Antigen of a Neisseria Meningitidis Vaccine by Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS). *Mol. Cell. Proteomics* **2024**, *23* (3), 100734. https://doi.org/10.1016/j.mcpro.2024.100734.

(40) Han, J.; Schmitz, A. J.; Richey, S. T.; Dai, Y.-N.; Turner, H. L.; Mohammed, B. M.; Fremont, D. H.; Ellebedy, A. H.; Ward, A. B. Polyclonal Epitope Mapping Reveals Temporal Dynamics and Diversity of Human Antibody Responses to H5N1 Vaccination. *Cell Rep.* **2021**, *34* (4), 108682. https://doi.org/10.1016/j.celrep.2020.108682.

(41) Nogal, B.; Bianchi, M.; Cottrell, C. A.; Kirchdoerfer, R. N.; Sewall, L. M.; Turner, H. L.; Zhao, F.; Sok, D.; Burton, D. R.; Hangartner, L.; Ward, A. B. Mapping Polyclonal Antibody Responses in Non-Human Primates Vaccinated with HIV Env Trimer Subunit Vaccines. *Cell Rep.* **2020**, *30* (11), 3755-3765.e7. https://doi.org/10.1016/j.celrep.2020.02.061.

(42) Ständer, S.; R. Grauslund, L.; Scarselli, M.; Norais, N.; Rand, K. Epitope Mapping of Polyclonal Antibodies by Hydrogen–Deuterium Exchange Mass Spectrometry (HDX-MS). *Anal. Chem.* **2021**, *93* (34), 11669–11678. https://doi.org/10.1021/acs.analchem.1c00696.

(43) Zhang, Q.; Noble, K. A.; Mao, Y.; Young, N. L.; Sathe, S. K.; Roux, K. H.; Marshall, A. G. Rapid Screening for Potential Epitopes Reactive with a Polycolonal Antibody by Solution-Phase H/D Exchange Monitored by FT-ICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (7), 1016–1025. https://doi.org/10.1007/s13361-013-0644-7.

(44) Vorauer, C.; Boniche-Alfaro, C.; Murphree, T.; Matsui, T.; Weiss, T.; Fries, B. C.; Guttman, M. Direct Mapping of Polyclonal Epitopes in Serum by HDX-MS. *Anal. Chem.* **2024**, *96* (42), 16758–16767. https://doi.org/10.1021/acs.analchem.4c03274.

(45) Robinson, C. V.; Sali, A.; Baumeister, W. The Molecular Sociology of the Cell. *Nature* **2007**, *450* (7172), 973–982. https://doi.org/10.1038/nature06523.

(46) Leung, M. R.; Zeng, J.; Wang, X.; Roelofs, M. C.; Huang, W.; Zenezini Chiozzi, R.; Hevler, J. F.; Heck, A. J. R.; Dutcher, S. K.; Brown, A.; Zhang, R.; Zeev-Ben-Mordehai, T. Structural Specializations of the Sperm Tail. *Cell* **2023**, *186* (13), 2880-2896.e17. https://doi.org/10.1016/j.cell.2023.05.026.

(47) Gui, M.; Wang, X.; Dutcher, S. K.; Brown, A.; Zhang, R. Ciliary Central Apparatus Structure Reveals Mechanisms of Microtubule Patterning. *Nat. Struct. Mol. Biol.* **2022**, *29* (5), 483–492. https://doi.org/10.1038/s41594-022-00770-2.

(48) Gui, M.; Ma, M.; Sze-Tu, E.; Wang, X.; Koh, F.; Zhong, E. D.; Berger, B.; Davis, J. H.; Dutcher, S. K.; Zhang, R.; Brown, A. Structures of Radial Spokes and Associated Complexes Important for Ciliary Motility. *Nat. Struct. Mol. Biol.* **2021**, *28* (1), 29–37. https://doi.org/10.1038/s41594-020-00530-0.

(49) Schmidt, L.; Tüting, C.; Kyrilis, F. L.; Hamdi, F.; Semchonok, D. A.; Hause, G.; Meister, A.; Ihling, C.; Stubbs, M. T.; Sinz, A.; Kastritis, P. L. Delineating Organizational Principles of the Endogenous L-A Virus by Cryo-EM and Computational Analysis of Native Cell Extracts. *Commun. Biol.* **2024**, *7* (1), 557. https://doi.org/10.1038/s42003-024-06204-7.

(50) Fianu, I.; Ochmann, M.; Walshe, J. L.; Dybkov, O.; Cruz, J. N.; Urlaub, H.; Cramer, P. Structural Basis of Integrator-Dependent RNA Polymerase II Termination. *Nature* **2024**, *629* (8010), 219–227. https://doi.org/10.1038/s41586-024-07269-4.

(51) Cingolani, G.; Iglesias, S.; Hou, C.-F.; Lemire, S.; Soriaga, A.; Kyme, P. Cryo-EM Analysis of Pseudomonas Phage Pa193 Structural Components. April 12, 2024. https://doi.org/10.21203/rs.3.rs-4189479/v1.

(52) Jiang, Y. X.; Cao, Q.; Sawaya, M. R.; Abskharon, R.; Ge, P.; DeTure, M.; Dickson, D. W.; Fu, J. Y.; Ogorzalek Loo, R. R.; Loo, J. A.; Eisenberg, D. S. Amyloid Fibrils in FTLD-TDP Are Composed of TMEM106B and Not TDP-43. *Nature* **2022**, *605* (7909), 304–309. https://doi.org/10.1038/s41586-022-04670-9.

(53) Hugener, J.; Xu, J.; Wettstein, R.; Ioannidi, L.; Velikov, D.; Wollweber, F.; Henggeler, A.; Matos, J.; Pilhofer, M. FilamentID Reveals the Composition and Function of Metabolic Enzyme Polymers during Gametogenesis. *Cell* **2024**, *187* (13), 3303-3318.e18. https://doi.org/10.1016/j.cell.2024.04.026.

(54) Ho, C.-M.; Li, X.; Lai, M.; Terwilliger, T. C.; Beck, J. R.; Wohlschlegel, J.; Goldberg, D. E.; Fitzpatrick, A. W. P.; Zhou, Z. H. Bottom-up Structural Proteomics: cryoEM of Protein Complexes Enriched from the Cellular Milieu. *Nat. Methods* **2020**, *17* (1), 79–85. https://doi.org/10.1038/s41592-019-0637-y.

(55) Chang, L.; Wang, F.; Connolly, K.; Meng, H.; Su, Z.; Cvirkaite-Krupovic, V.; Krupovic, M.; Egelman, E. H.; Si, D. DeepTracer-ID: De Novo Protein Identification from Cryo-EM Maps. *Biophys. J.* **2022**, *121* (15), 2840–2848. https://doi.org/10.1016/j.bpj.2022.06.025.

(56) Chojnowski, G.; Simpkin, A. J.; Leonardo, D. A.; Seifert-Davila, W.; Vivas-Ruiz, D. E.; Keegan, R. M.; Rigden, D. J. *findMySequence*: A Neural-Network-Based Approach for Identification of Unknown Proteins in X-Ray Crystallography and Cryo-EM. *IUCrJ* **2022**, *9* (1), 86–97. https://doi.org/10.1107/S2052252521011088.

(57) Jamali, K.; Käll, L.; Zhang, R.; Brown, A.; Kimanius, D.; Scheres, S. H. W. Automated Model Building and Protein Identification in Cryo-EM Maps. *Nature* **2024**, *628* (8007), 450–457. https://doi.org/10.1038/s41586-024-07215-4.

(58) Schulte, D.; Peng, W.; Snijder, J. Template-Based Assembly of Proteomic Short Reads For *De Novo* Antibody Sequencing and Repertoire Profiling. *Anal. Chem.* **2022**, *94* (29), 10391–10399. https://doi.org/10.1021/acs.analchem.2c01300.

(59) Schulte, D.; Snijder, J. A Handle on Mass Coincidence Errors in *de Novo* Sequencing of Antibodies by Bottom-up Proteomics. February 22, 2024. https://doi.org/10.1101/2024.02.20.581155.

(60) Manso, T.; Folch, G.; Giudicelli, V.; Jabado-Michaloud, J.; Kushwaha, A.; Nguefack Ngoune, V.; Georga, M.; Papadaki, A.; Debbagh, C.; Pégorier, P.; Bertignac, M.; Hadi-Saljoqi, S.; Chentli, I.; Cherouali, K.; Aouinti, S.; El Hamwi, A.; Albani, A.; Elazami Elhassani, M.; Viart, B.; Goret, A.; Tran, A.; Sanou, G.; Rollin, M.; Duroux, P.; Kossida, S. IMGT® Databases, Related Tools and Web Resources through Three Main Axes of Research and Development. *Nucleic Acids Res.* **2022**, *50* (D1), D1262–D1272. https://doi.org/10.1093/nar/gkab1136.

(61) Lefranc, M.-P. Immunoglobulin and T Cell Receptor Genes: IMGT® and the Birth and Rise of Immunoinformatics. *Front. Immunol.* **2014**, *5*. https://doi.org/10.3389/fimmu.2014.00022.
(62) Lefranc, M.-P.; Giudicelli, V.; Duroux, P.; Jabado-Michaloud, J.; Folch, G.; Aouinti, S.; Carillon, E.; Duvergey, H.; Houles, A.; Paysan-Lafosse, T.; Hadi-Saljoqi, S.; Sasorith, S.; Lefranc, G.; Kossida, S. IMGT®, the International ImMunoGeneTics Information System® 25 Years On. *Nucleic Acids Res.* **2015**, *43* (D1), D413–D422. https://doi.org/10.1093/nar/gku1056.
(63) Momont, C.; Dang, H. V.; Zatta, F.; Hauser, K.; Wang, C.; Di Iulio, J.; Minola, A.; Czudnochowski, N.; De Marco, A.; Branch, K.; Donermeyer, D.; Vyas, S.; Chen, A.; Ferri, E.; Guarino, B.; Powell, A. E.; Spreafico, R.; Yim, S. S.; Balce, D. R.; Bartha, I.; Meury, M.; Croll, T. I.; Belnap, D. M.; Schmid, M. A.; Schaiff, W. T.; Miller, J. L.; Cameroni, E.; Telenti, A.; Virgin, H. W.; Rosen, L. E.; Purcell, L. A.; Lanzavecchia, A.; Snell, G.; Corti, D.; Pizzuto, M. S. A Pan-Influenza Antibody Inhibiting Neuraminidase via Receptor Mimicry. *Nature* **2023**, *618* (7965), 590–597. https://doi.org/10.1038/s41586-023-06136-y.
(64) Ter Meulen, J.; Van Den Brink, E. N.; Poon, L. L. M.; Marissen, W. E.; Leung, C. S. W.; Cox, F.; Cheung, C. Y.; Bakker, A. Q.; Bogaards, J. A.; Van Deventer, E.; Preiser, W.; Doerr, H. W.; Chow, V. T.; De Kruif, J.; Peiris, J. S. M.; Goudsmit, J. Human Monoclonal Antibody Combination against SARS Coronavirus: Synergy and Coverage of Escape Mutants. *PLoS Med.* **2006**, *3* (7), e237. https://doi.org/10.1371/journal.pmed.0030237.
(65) Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C.-C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A. A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368* (6491), 630–633. https://doi.org/10.1126/science.abb7269.
(66) Wrobel, A. G.; Benton, D. J.; Hussain, S.; Harvey, R.; Martin, S. R.; Roustan, C.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Antibody-Mediated Disruption of the SARS-CoV-2 Spike Glycoprotein. *Nat. Commun.* **2020**, *11* (1), 5337. https://doi.org/10.1038/s41467-020-19146-5.
(67) Huo, J.; Zhao, Y.; Ren, J.; Zhou, D.; Duyvesteyn, H. M. E.; Ginn, H. M.; Carrique, L.; Malinauskas, T.; Ruza, R. R.; Shah, P. N. M.; Tan, T. K.; Rijal, P.; Coombes, N.; Bewley, K. R.; Tree, J. A.; Radecke, J.; Paterson, N. G.; Supasa, P.; Mongkolsapaya, J.; Screaton, G. R.; Carroll, M.; Townsend, A.; Fry, E. E.; Owens, R. J.; Stuart, D. I. Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host Microbe* **2020**, *28* (3), 445-454.e6. https://doi.org/10.1016/j.chom.2020.06.010.
(68) Gadush, M. V.; Sautto, G. A.; Chandrasekaran, H.; Bensussan, A.; Ross, T. M.; Ippolito, G. C.; Person, M. D. Template-Assisted *De Novo* Sequencing of SARS-CoV-2 and Influenza Monoclonal Antibodies by Mass Spectrometry. *J. Proteome Res.* **2022**, *21* (7), 1616–1627. https://doi.org/10.1021/acs.jproteome.1c00913.
(69) Ferguson, J. A.; Raghavan, S. S. R.; Peña Alzua, G.; Bhavsar, D.; Huang, J.; Rodriguez, A. J.; Torres, J. L.; Bottermann, M.; Han, J.; Krammer, F.; Batista, F. D.; Ward, A. B. Functional and Epitope Specific Monoclonal Antibody Discovery Directly from Immune Sera Using cryoEM. Immunology December 9, 2024. https://doi.org/10.1101/2024.12.06.627063.
(70) Mistry, J.; Finn, R. D.; Eddy, S. R.; Bateman, A.; Punta, M. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. *Nucleic Acids Res.* **2013**, *41* (12), e121–e121. https://doi.org/10.1093/nar/gkt263.
(71) Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P. C.; Timony, M. A.; Ottaviano, M.; Sliz, P. Collaboration Gets the Most out of Software. *eLife* **2013**, *2*, e01456. https://doi.org/10.7554/eLife.01456.