

Mass Spectrometry-Based *De Novo* Sequencing of Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme

Weiwei Peng,[§] Matti F. Pronker,[§] and Joost Snijder*Cite This: *J. Proteome Res.* 2021, 20, 3559–3566

Read Online

ACCESS |



Metrics & More



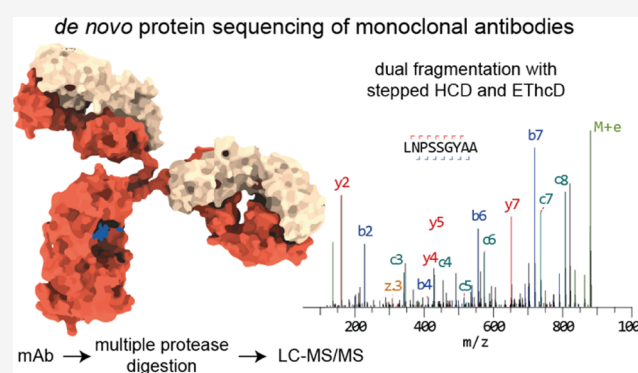
Article Recommendations



Supporting Information

ABSTRACT: Antibody sequence information is crucial to understanding the structural basis for antigen binding and enables the use of antibodies as therapeutics and research tools. Here, we demonstrate a method for direct *de novo* sequencing of monoclonal IgG from the purified antibody products. The method uses a panel of multiple complementary proteases to generate suitable peptides for *de novo* sequencing by liquid chromatography–tandem mass spectrometry (LC-MS/MS) in a bottom-up fashion. Furthermore, we apply a dual fragmentation scheme, using both stepped high-energy collision dissociation (stepped HCD) and electron-transfer high-energy collision dissociation (EThcD), on all peptide precursors. The method achieves full sequence coverage of the monoclonal antibody herceptin, with an accuracy of 99% in the variable regions. We applied the method to sequence the widely used anti-FLAG-M2 mouse monoclonal antibody, which we successfully validated by remodeling a high-resolution crystal structure of the Fab and demonstrating binding to a FLAG-tagged target protein in Western blot analysis. The method thus offers robust and reliable sequences of monoclonal antibodies.

KEYWORDS: mass spectrometry, antibody, *de novo* sequencing, EThcD, stepped HCD, herceptin, FLAG-tag, anti-FLAG-M2



INTRODUCTION

Antibodies can bind a great molecular diversity of antigens, owing to the high degree of sequence diversity that is available through somatic recombination, hypermutation, and heavy–light-chain pairings.^{1,2} Sequence information on antibodies therefore is crucial to understanding the structural basis of antigen binding, how somatic hypermutation governs affinity maturation, and an overall understanding of the adaptive immune response in health and disease, by mapping out the antibody repertoire. Moreover, antibodies have become invaluable research tools in the life sciences and ever more widely developed as therapeutic agents.^{3,4} In this context, sequence information is crucial for the use, production, and validation of these important research tools and biopharmaceutical agents.^{5,6}

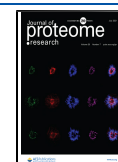
Antibody sequences are typically obtained through cloning and sequencing of the coding mRNAs of the paired heavy and light chains.^{7–9} The sequencing workflows thereby rely on isolation of the antibody-producing cells from peripheral blood monocytes, or spleen and bone marrow tissues. These antibody-producing cells are not always readily available, however, and cloning/sequencing of the paired heavy and light chains is a nontrivial task with a limited success rate.^{7–9} Moreover, antibodies are secreted in bodily fluids and mucus. Antibodies are thereby in large part functionally disconnected

from their producing B-cell, which raises questions on how the secreted antibody pool relates quantitatively to the underlying B-cell population and whether there are potential sampling biases in current antibody sequencing strategies.

Direct mass spectrometry (MS)-based sequencing of the secreted antibody products is a useful complementary tool that can address some of the challenges faced by conventional sequencing strategies relying on cloning/sequencing of the coding mRNAs.^{10–17} MS-based methods do not rely on the availability of the antibody-producing cells but rather target the polypeptide products directly, offering the prospect of a next generation of serology, in which secreted antibody sequences might be obtained from any bodily fluid. Whereas MS-based *de novo* sequencing still has a long way to go toward this goal, owing to limitations in sample requirements, sequencing accuracy, read length, and sequence assembly, MS has been successfully used to profile the antibody repertoire and obtain

Received: March 1, 2021

Published: June 14, 2021



(partial) antibody sequences beyond those available from conventional sequencing strategies based on cloning/sequencing of the coding mRNAs.^{10–17}

Most MS-based strategies for antibody sequencing rely on a proteomics-type bottom-up liquid chromatography-tandem mass spectrometry (LC-MS/MS) workflow, in which the antibody product is digested into smaller peptides for MS analysis.^{14,18–23} Available germline antibody sequences are then often used either as a template to guide assembly of *de novo* peptide reads (such as in PEAKS Ab)²³ or as a starting point to iteratively identify somatic mutations to arrive at the mature antibody sequence (such as in Supernovo).²¹ To maximize sequence coverage and aid read assembly, these MS-based workflows typically use a combination of complementary proteases and unspecific digestion to generate overlapping peptides. The most straightforward application of these MS-based sequencing workflows is the successful sequencing of monoclonal antibodies from (lost) hybridoma cell lines but it also forms the basis of more advanced and challenging applications to characterize polyclonal antibody mixtures and profile the full antibody repertoire from serum.

Here, we describe an efficient protocol for MS-based sequencing of monoclonal antibodies. The protocol requires approximately 200 pmol of the antibody product, and sample preparation can be completed within 1 working day. We selected a panel of nine proteases with complementary specificities, which are active in the same buffer conditions for parallel digestion of the antibodies. We developed a dual fragmentation strategy for MS/MS analysis of the resulting peptides to yield rich sequence information from the fragmentation spectra of the peptides. The protocol yields full and deep sequence coverages of the variable domains of both heavy and light chains, as demonstrated on the monoclonal antibody herceptin. As a test case, we used our protocol to sequence the widely used anti-FLAG-M2 mouse monoclonal antibody, for which no sequence was publicly available despite its described use in 5000+ peer-reviewed publications.^{24,25} The protocol achieved full sequence coverage of the variable domains of both heavy and light chains, including all complementarity determining regions (CDRs). The obtained sequence was successfully validated by remodeling the published crystal structure of the anti-FLAG-M2 Fab and demonstrating binding of the synthetic recombinant antibody following the experimental sequence to a FLAG-tagged protein in Western blot analysis. The protocol developed here thus offers robust and reliable sequencing of monoclonal antibodies with prospective applications for sequencing secreted antibodies from bodily fluids.

■ EXPERIMENTAL SECTION

Sample Preparation

Anti-FLAG M2 antibody was purchased from Sigma-Aldrich (catalogue number F1804). Herceptin was provided by Roche (Penzberg, Germany). Twenty-seven micrograms of each sample was denatured in 2% sodium deoxycholate (SDC), 200 mM Tris-HCl, 10 mM Tris(2-carboxyethyl)phosphine (TCEP), pH 8.0 at 95 °C for 10 min, followed by 30 min incubation at 37 °C for reduction. The sample was then alkylated by adding iodoacetic acid to a final concentration of 40 mM and incubated in the dark at room temperature for 45 min. A 3 μg sample was then digested by one of the following

proteases: trypsin, chymotrypsin, lysN, lysC, gluC, aspN, aLP, thermolysin, and elastase in a 1:50 ratio (w/w) in a total volume of 100 μL of 50 mM ammonium bicarbonate at 37 °C for 4 h. After digestion, SDC was removed by adding 2 μL of formic acid (FA) and centrifugation at 14 000g for 20 min. Following centrifugation, the supernatant containing the peptides was collected for desalting on a 30 μm Oasis HLB 96-well plate (Waters). The Oasis HLB sorbent was activated with 100% acetonitrile and subsequently equilibrated with 10% formic acid in water. Next, peptides were bound to the sorbent, washed twice with 10% formic acid in water, and eluted with 100 μL of 50% acetonitrile/5% formic acid in water (v/v). The eluted peptides were vacuum-dried and reconstituted in 100 μL of 2% FA.

Mass Spectrometry

The digested peptides (single injection of 0.2 μg) were separated by online reversed phase chromatography on an Agilent 1290 UHPLC (column packed with Poroshell 120 EC C18; dimensions 50 cm × 75 μm, 2.7 μm, Agilent Technologies) coupled to a Thermo Scientific Orbitrap Fusion mass spectrometer. 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) as well as electron-transfer high-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.

MS Data Analysis

Automated *de novo* sequencing was performed with Supernovo (version 3.10, Protein Metrics Inc.). Supernovo identifies closely matching antibody germline sequences from an initial database search, followed by iterative substitutions to the recombined V–J–C sequences of heavy and light chains by wildcard searches on the MS/MS spectra to converge to the final output sequence.²¹ Custom parameters were used as follows: nonspecific digestion; precursor and product mass tolerance were set to 12 ppm and 0.02 Da, respectively; carboxymethylation (+58.005479) on cysteine was set as fixed modification; oxidation on methionine and tryptophan was set as variable common 1 modification; carboxymethylation on the N-terminus, pyroglutamic acid conversion of glutamine and glutamic acid on the N-terminus, and deamidation on asparagine/glutamine were set as variable rare 1 modifications. Peptides were filtered for a score ≥500 for the final evaluation of spectrum quality and (depth of) coverage. The “depth of coverage” was defined as the number of unique peptides with a score ≥500 that cover the position. Supernovo generates peptide groups for redundant MS/MS spectra, including also when both stepped HCD and EThcD fragmentation on the same precursor generate good peptide-spectrum matches. In these cases, only the best-matched spectrum is counted as representative for that group. This criterion was used in

counting the number of peptide reads reported in Table S1. Germline sequences and CDR boundaries were inferred using IMGT/DomainGapAlign.^{26,27}

Revision of the Anti-FLAG-M2 Fab Crystal Structure Model

As a starting point for model building, the reflection file and coordinates of the published anti-FLAG-M2 Fab crystal structure were used (PDB ID: 2G60).²⁸ Care was taken to use the original R_{free} labels of the deposited reflection file for refinement, so as not to introduce extra model bias. Differential residues between this structure and our mass spectrometry-derived anti-FLAG sequence were manually mutated and fitted in the density using Coot.²⁹ Many spurious water molecules that caused severe steric clashes in the original model were also manually removed in Coot. Densities for two sulfate and one chloride ion were identified and built into the model. The original crystallization solution contained 0.1 M ammonium sulfate. Iterative cycles of model geometry optimization in real space in Coot and reciprocal space refinement by Phenix were used to generate the final model, which was validated with Molprobtly.^{30,31}

Cloning and Expression of Synthetic Recombinant Anti-FLAG-M2

To recombinantly express full-length anti-FLAG-M2, the proteomic sequences of both the light and heavy chains were reverse-translated and codon-optimized for expression in human cells using the Integrated DNA Technologies (IDT) web tool (<http://www.idtdna.com/CodonOpt>).³² For the linker and Fc region of the heavy chain, the standard mouse Ig γ -1 (IGHG1) amino acid sequence (Uniprot P01868.1) was used. An N-terminal secretion signal peptide derived from human IgG light chain (MEAPAQLLFLLLWLPDPTTG) was added to the N-termini of both heavy and light chains. BamHI and NotI restriction sites were added to the 5' and 3' ends of the coding regions, respectively. Only for the light chain, a double stop codon was introduced at the 3' site before the NotI restriction site. The coding regions were subcloned using BamHI and NotI restriction-ligation into a pRK5 expression vector with a C-terminal octahistidine tag between the NotI site and a double stop codon 3' of the insert, so that only the heavy chain has a C-terminal AAAHHHHHHHH sequence for nickel-affinity purification (the triple alanine resulting from the NotI site). The L51I correction in the heavy chain was introduced later (after observing it in the crystal structure) by *in vivo* assembly (IVA) cloning.³³ Expression plasmids for the heavy and light chains were mixed in a 1:1 (w/w) ratio for transient transfection in HEK293 cells with polyethylenimine, following standard procedures. The medium was collected 6 days after transfection, and cells were spun down by 10 min of centrifugation at 1000g. The antibody was directly purified from the supernatant using Ni-sepharose excel resin (Cytiva Life Sciences), washing with 500 mM NaCl, 2 mM CaCl₂, 15 mM imidazole, 20 mM N-(2-hydroxyethyl)piperazine-N'-ethanesulfonic acid (HEPES) pH 7.8 and eluting with 500 mM NaCl, 2 mM CaCl₂, 200 mM imidazole, 20 mM HEPES pH 7.8.

Western Blot Validation of Anti-FLAG-M2 Binding

To test the binding of our recombinant anti-FLAG-M2 to the FLAG-tag epitope, compared to the commercially available anti-FLAG-M2 (Sigma-Aldrich), we used both antibodies to probe Western blots of a FLAG-tagged protein in parallel. Purified Rabies virus glycoprotein ectodomain (SAD B19

strain, UNIPROT residues 20–450) with or without a C-terminal FLAG-tag followed by a foldon trimerization domain and an octahistidine tag was heated to 95 °C in XT sample buffer (Bio-Rad) for 5 min. Samples were run twice on a Criterion XT 4–12% polyacrylamide gel (Bio-Rad) in MES XT buffer (Bio-Rad) before Western blot transfer to a nitrocellulose membrane in Tris–glycine buffer (Bio-Rad) with 20% methanol. The membrane was blocked with 5% (w/v) dry nonfat milk in phosphate-buffered saline (PBS) overnight at 4 °C. The membrane was cut into two (one half for the commercial and one half for the recombinant anti-FLAG-M2), and each half was probed with either commercial (Sigma-Aldrich) or recombinant anti-FLAG-M2 at 1 $\mu\text{g}/\text{mL}$ in PBS for 45 min. After washing three times with PBST (PBS with 0.1% v/v Tween20), polyclonal goat antimouse fused to horseradish peroxidase (HRP) was used to detect binding of anti-FLAG-M2 to the FLAG-tagged protein for both membranes. The membranes were washed three more times with PBST before applying enhanced chemiluminescence (ECL; Pierce) reagent to image the blots in parallel.

RESULTS

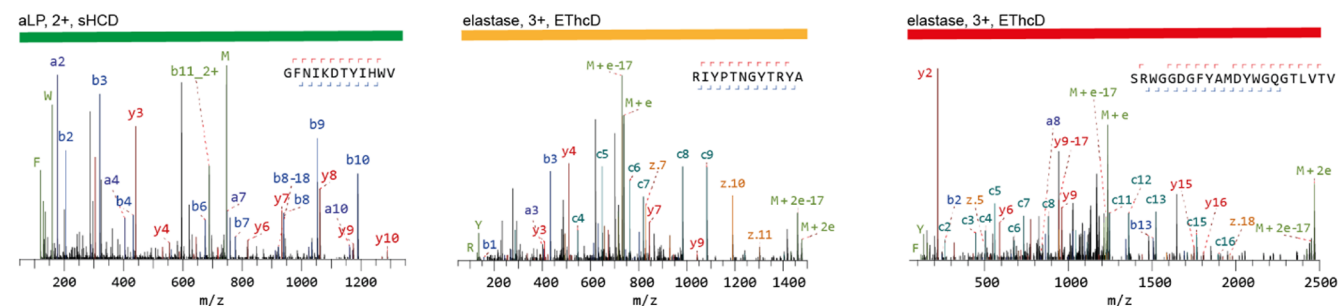
We used an in-solution digestion protocol, with sodium deoxycholate as the denaturing agent, to generate peptides from the antibodies for LC-MS/MS analysis. Following heat denaturation and disulfide bond reduction, we used iodoacetic acid as the alkylating agent to cap free cysteines. Note that conventional alkylating agents like iodo-/chloroacetamide generate +57 Da mass differences on cysteines and primary amines, which may lead to spurious assignments as glycine residues in *de novo* sequencing. The +58 Da mass differences generated by alkylation with iodoacetic acid circumvents this potential pitfall.

We chose a panel of nine proteases with activity at pH 7.5–8.5, so that the denatured, reduced, and alkylated antibodies could be easily split for parallel digestion under the same buffer conditions. These proteases (with indicated cleavage specificities) included trypsin (C-terminal of R/K), chymotrypsin (C-terminal of F/Y/W/M/L), α -lytic protease (C-terminal of T/A/S/V), elastase (unspecific), thermolysin (unspecific), lysN (N-terminal of K), lysC (C-terminal of K), aspN (N-terminal of D/E), and gluC (C-terminal of D/E). Correct placement or assembly of peptide reads is a common challenge in *de novo* sequencing, which can be facilitated by sufficient overlap between the peptide reads. This favors the occurrence of missed cleavages and longer reads, so we opted to perform a brief 4 h digestion. Following digestion, SDC is removed by precipitation and the peptide supernatant is desalted, ready for LC-MS/MS analysis. The resulting raw data were used for automated *de novo* sequencing with the Supernovo software package.

As peptide fragmentation is dependent on many factors like length, charge state, composition, and sequence,³⁴ we needed a versatile fragmentation strategy to accommodate the diversity of antibody-derived peptides generated by the nine proteases. We opted for a dual fragmentation scheme that applies to both stepped high-energy collision dissociation (stepped HCD) and electron-transfer high-energy collision dissociation (ETHcD) on all peptide precursors.^{35–37} The stepped HCD fragmentation includes three collision energies to cover multiple dissociation regimes, and the ETHcD fragmentation works especially well for higher charged states, also adding complementary *c/z* ions for maximum sequence coverage.

(A) Heavy Chain

de novo | EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSLRRAEDTAVYYCSRWGGDGFYAMDYWGQGLTIVTSS
 herceptin | EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSLRRAEDTAVYYCSRWGGDGFYAMDYWGQGLTIVTSS



(B) Light Chain

de novo | DIQMTQSPSSLSASVGRVITTCRASQDVTAVAWYQKPGKAPKLLIYSASFLYSGVSPRFSGSRSGTDFTLTLSSLPEDFATYYCQQWNTTPTTFFGGQTKVEIKRTV
 herceptin | DIQMTQSPSSLSASVGRVITTCRASQDVTAVAWYQKPGKAPKLLIYSASFLYSGVSPRFSGSRSGTDFTLTLSSLPEDFATYYCQQHYTTPPTFFGGQTKVEIKRTV

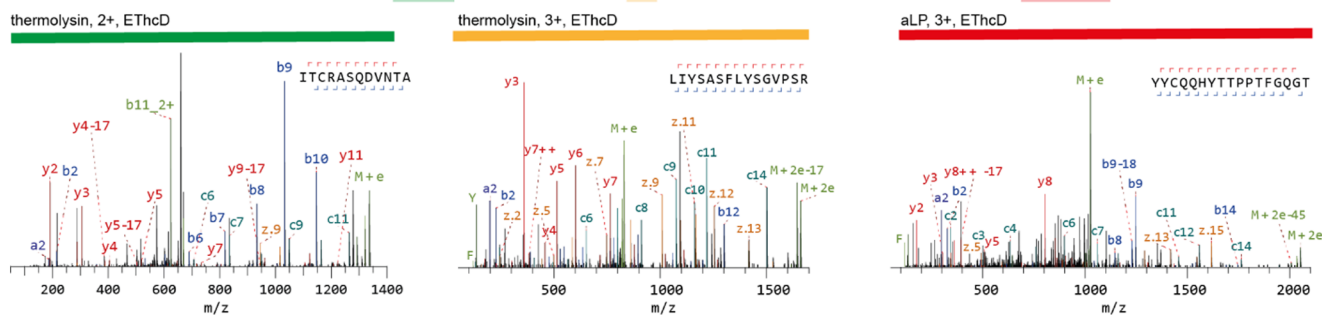


Figure 1. Mass spectrometry-based *de novo* sequencing of the monoclonal antibody herceptin. The variable regions of the heavy (A) and light chains (B) are shown. The MS-based sequence is shown alongside the known herceptin sequence, with differences highlighted by asterisks (*). Exemplary MS/MS spectra supporting the assigned sequences of the heavy- and light-chain CDRs are shown below the alignments with protease, precursor charge state, and fragmentation type indicated. Peptide sequence and fragment coverage are indicated on top of the spectra, with *b/c* ions indicated in blue/teal and *y/z* ions in red/orange. The same coloring is used to annotate peaks in the spectra, with additional peaks such as intact/charge reduced precursors, neutral losses, and immonium ions indicated in green. Note that to prevent overlapping peak labels, only a subset of successfully matched peaks are annotated.

We used the monoclonal antibody herceptin (also known as trastuzumab) as a benchmark to test our protocol.^{38,39} From the total data set of nine proteases, we collected 4408 peptide reads (defined as peptides with a score ≥ 500 ; see [Experimental Section](#) for details), of which 2866 are with superior stepped HCD fragmentation (compared to EThcD) and conversely 1722 peptide reads are with superior EThcD fragmentation (see [Table S1](#)). Sequence coverage was 100% in both heavy and light chains across the variable and constant domains (see [Figures S1 and S2](#)). The median depth of coverage was 148 overall and slightly higher in the light chain (see [Table S1 and Figures S1 and S2](#)). The median depth of coverage in the CDRs of both chains ranged from 42 to 210.

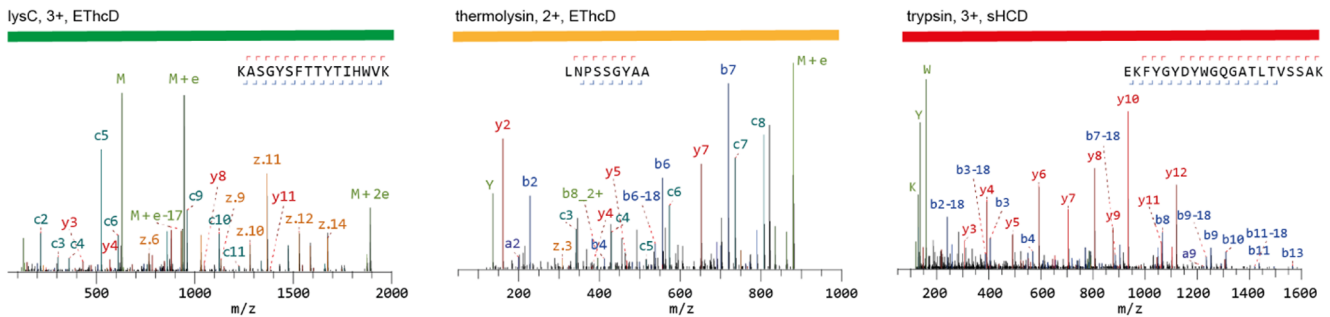
The experimentally determined *de novo* sequence is shown alongside the known herceptin sequence for the variable domains of both chains in [Figure 1](#), with exemplary MS/MS spectra for the CDRs. We achieved an overall sequence accuracy of 99% with the automated sequencing procedure of Supernovo, with three incorrect assignments in the light chain. In framework 3 of the light chain, I75 was incorrectly assigned as the isomer leucine (L), a common MS-based sequencing error. In CDRL3 of the light chain, an additional misassignment was made for the dipeptide H91/Y92, which was incorrectly assigned as W91/N92. The dipeptides HY and WN have identical masses, and the misassignment of W91/N92 (especially W91) was poorly supported by the fragmentation

spectra, in contrast to the correct H91/Y92 assignment (see *c6/c7* in fragmentation spectra; [Figure 1](#)). Overall, the protocol yielded highly accurate sequences at a combined 230/233 position of the variable domains in herceptin.

The combined use of both stepped HCD and EThcD fragmentation resulted in superior accuracy compared to the separate fragmentation techniques (see [Figure S3](#)). Likewise, the use of all nine proteases resulted in superior accuracy compared to a smaller subset of trypsin, chymotrypsin, thermolysin, and elastase or single protease data sets (see [Figure S3](#)). Finally, compared to overnight digestion, the shorter 4 h digestion of our protocol resulted in peptides of similar lengths (see [Figure S4](#)). However, specific proteases showed different effects of digestion time; overnight digestion gave a higher number of peptides for trypsin and chymotrypsin but fewer for elastase and thermolysin digestion. From the subset of four proteases (trypsin, chymotrypsin, elastase, and thermolysin) used for this comparison, the overnight digestion resulted in fewer errors compared to the 4 h digestion overall, to an equivalent number as observed in the full data set with nine proteases. The main benefit of the shortened digestion is therefore that the sample preparation can be completed within 1 working day. These comparisons highlight that the key to accurate sequencing with our protocol is the dual fragmentation scheme in combination with the multitude of proteases for antibody digestion, rather than digestion time, and that the

(A) Heavy Chain

		CDRH1		CDRH2		CDRH3
de novo		QVQLQQAELARPGASVKMSCKASGYSFTTYTIHWVKRPGQGLEWIGYLNPSGGYAAYNQNFKDETTLTADPSSSTAYMELNSLTSEDSAVVYCAREKFGYDYWGQGATLTVSS				
crystal		EVQLQSSGGELAKPGASVKMSCKSSGYTFYAIHWKQAAGAGLEWIGYIAPAAGAAAYNAAFKFKATLAADKSSSTAYMAAAALTSSEDSAVVYCARAAAAGADYWGQGATLTVSS				
germline		QVQLQQAELARPGASVKMSCKASGYSFTSYTMHWVKRPGQGLEWIGYINPSGGYTEYNQKFKDKTTLTADKSSSTAYMQLSSLTSEDSAVVYCAR----YFDYWGQGATLTVSS				
		* * *	* * * * * *	* * * * * * * * *	* * * * *	* * * * * * *



(B) Light Chain

		CDRL1		CDRL2		CDRL3
de novo		DVLMTQIPLSLPVSLGDAQISCRSSQIVHRNGNTYLEWYLLKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYYCFQGSHPVPTFGGGTKLEI				
crystal		DVLMTQAPLTLVSLGDAQISCRSSQAIHVHANGNTYLEWYLLKPGQSPALLIYKVANRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYYCFQGAHAPYTFGGGTKLEI				
germline		DVLMTQTPSLPVSLGDAQISCRSSQIVHSNGNTYLEWY-KPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYYCFQGSHPVPTFGGGTKLEI				
		* *	* *	* *	* *	* *

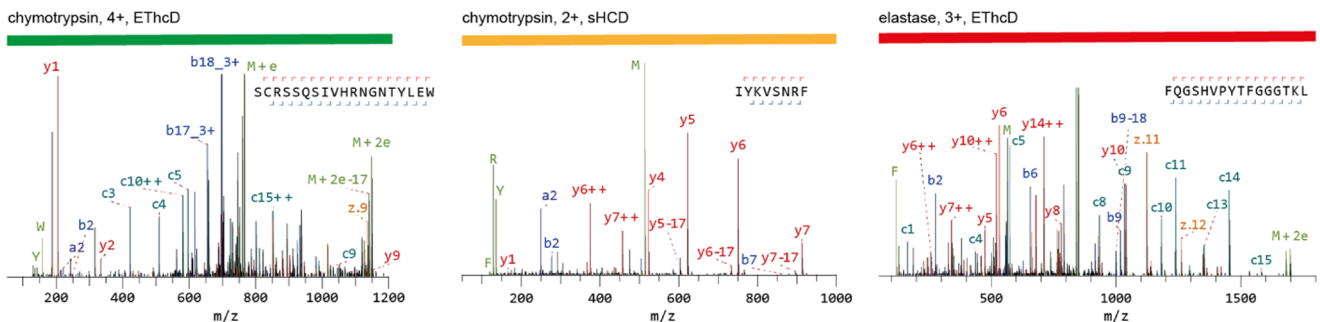


Figure 2. Mass spectrometry-based *de novo* sequence of the mouse monoclonal anti-FLAG-M2 antibody. The variable regions of the heavy (A) and light chains (B) are shown. The MS-based sequence is shown alongside the previously published sequence in the crystal structure of the Fab (PDB ID: 2G60) and germline sequence (IMGT-DomainGapAlign; IGHV1-04/IGHJ2; IGKV1-117/IGKJ1). Differential residues are highlighted by asterisks (*). Exemplary MS/MS spectra in support of the assigned sequences are shown below the alignments with protease, precursor charge state, and fragmentation type indicated. Peptide sequence and fragment coverage are indicated on top of the spectra, with *b/c* ions indicated in blue/teal and *y/z* ions in red/orange. The same coloring is used to annotate peaks in the spectra, with additional peaks such as intact/charge reduced precursors, neutral losses, and immonium ions indicated in green. Note that to prevent overlapping peak labels, only a subset of successfully matched peaks is annotated.

protocol could be further optimized by adapting digestion time for specific proteases individually.

We next applied our sequencing protocol to the mouse monoclonal anti-FLAG-M2 antibody as a test case.²⁴ Despite the widespread use of anti-FLAG-M2 to detect and purify FLAG-tagged proteins,⁴⁰ the only publicly available sequences can be found in the crystal structure of the Fab.²⁸ The modeled sequence of the original crystal structure had to be inferred from germline sequences that could match the experimental electron density and also includes many placeholder alanines at positions that could not be straightforwardly interpreted. The full anti-FLAG-M2 data set from the nine proteases included 3371 peptide reads (with scores ≥ 500): 1983 with superior stepped HCD fragmentation spectra (compared to EThcD) and conversely 1388 with superior EThcD spectra. We achieved full sequence coverage of the variable regions of both heavy and light chains, with a median depth of coverage in the CDRs ranging from 32 to 192 (see Table S1). As for herceptin, the depth of coverage was better in the light chain compared to that in the heavy chain (see Figures S1 and S2). The full MS-based anti-FLAG-M2 sequences can be found in FASTA format in the Supporting Information.

The MS-based sequences of anti-FLAG-M2 are shown alongside the crystal structure sequences and the inferred germline precursors with exemplary MS/MS spectra for the CDRs in Figure 2. The experimentally determined sequence reveals that anti-FLAG-M2 is a mouse IgG1, with an IGHV1-04/IGHJ2 heavy chain and IGKV1-117/IGKJ1 kappa light chain. The experimentally determined sequence differs at 34 and 9 positions in the heavy and light chains of the Fab crystal structure, respectively. To validate the experimentally determined sequences, we remodeled the crystal structure using the MS-based heavy and light chains, resulting in much improved model statistics (see Figure 3 and Table S2). The experimental electron densities show excellent support of the MS-based sequence (as shown for the CDRs in Figure 3B). A notable exception is L51 in CDRH2 of the heavy chain. The MS-based sequence was assigned as leucine, but the experimental electron density supports assignment of the isomer isoleucine instead (see Figure S5). In contrast to the original model, our new MS-based model reveals a predominantly positively charged paratope (see Figure S6), which potentially complements the -3 net charge of the FLAG-tag epitope (DYKDDDDK) to mediate binding. The experimentally determined anti-FLAG-M2 sequence, with the

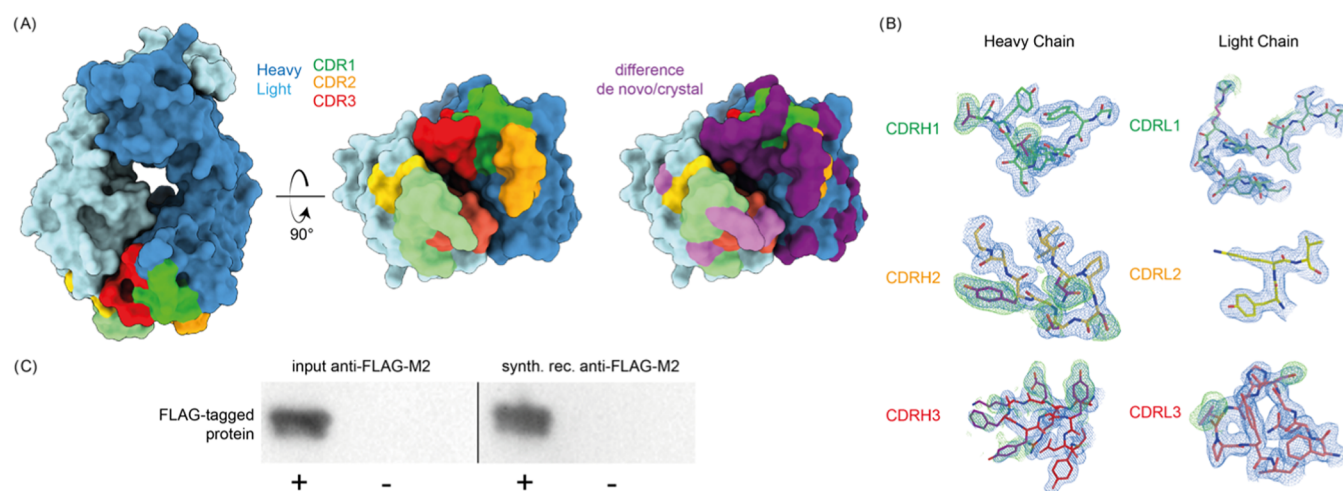


Figure 3. Validation of the MS-based anti-FLAG-M2 sequence. (A) Previously published crystal structure of the anti-FLAG-M2 Fab was remodeled with the experimentally determined sequence, shown in surface rendering with CDRs and differential residues highlighted in colors. (B) $2F_o - F_c$ electron density of the newly refined map contoured at 1 root-mean-square deviation (RMSD) is shown in blue, and $F_o - F_c$ positive difference density of the original deposited map contoured at 1.7 RMSD is shown in green around the CDR loops of the heavy and light chains. Differential residues between the published crystal structure and the model based on our antibody sequencing are indicated in purple. (C) Western blot validation of the synthetic recombinant anti-FLAG-M2 antibody produced with the experimentally determined sequence demonstrates equivalent FLAG-tag binding compared to commercial anti-FLAG-M2 (see also Figure S3).

LSII correction, was further validated by testing binding of the synthetic recombinant antibody to a purified FLAG-tagged protein in Western blot analysis (see Figures 3C and S7). The synthetic recombinant antibody showed equivalent binding compared to the original antibody sample used for sequencing, confirming that the experimentally determined sequence is reliable to obtain the recombinant antibody product with the desired functional profile.

DISCUSSION

There are four other monoclonal antibody sequences against the FLAG-tag publicly available through the AntiBodies Chemically Defined (ABCD) database.^{41–43} Comparison of the CDRs of anti-FLAG-M2 with these additional four monoclonal antibodies reveals a few common motifs that may determine FLAG-tag binding specificity (see Table S3). In the heavy chain, the only common motif between all five monoclonals is that the first three residues of CDRH1 follow a GXS sequence. In addition, the last three residues of CDRH3 of anti-FLAG-M2 are YDY, similar to MDY in 2H8 and YDF in EEh13.6 (and EEh14.3 also ends CDRH3 with an aromatic F residue). In contrast to the heavy chain, the CDRs of the light chain are almost completely conserved in 4/5 monoclonals with only minimal differences compared to germline. The anti-FLAG-M2 and 2H8 monoclonals were specifically raised in mice against the FLAG-tag epitope,^{24,42} whereas the computationally designed EEh13.6 and EEh14.3 monoclonals contain the same light chain from an EE-dipeptide tag-directed antibody.⁴¹ This suggests that the IGKV1-117/IGKJ1 light chain may be a common determinant of binding to a small negatively charged peptide epitope like the FLAG-tag and is readily available as a hardcoded germline sequence in the mouse antibody repertoire.

The availability of the anti-FLAG-M2 sequences may contribute to the wider use of this important research tool, as well as the development and engineering of better FLAG-tag-directed antibodies. This example illustrates that our MS-based sequencing protocol yields robust and reliable

monoclonal antibody sequences. The protocol described here also formed the basis of a recent application where we sequenced an antibody directly from patient-derived serum, using a combination with top-down fragmentation of the isolated Fab fragment.⁴⁴ The dual fragmentation strategy yields high-quality spectra suitable for *de novo* sequencing and may further contribute to the exciting prospect of a new advanced serology in which antibody sequences can be directly obtained from bodily fluids.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00169>.

Anti-FLAG-M2 MS-based sequences in FASTA format; coverage statistics for the herceptin benchmark and anti-FLAG-M2 MAb sequences (Table S1); model statistics for Fab crystal structure (Table S2); comparison of CDR sequences of FLAG-tag binding MABs (Table S3); coverage maps of anti-FLAG-M2 MAB sequences (Figure S1); depth of coverage profiles for herceptin and anti-FLAG-M2 sequences (Figure S2); sequence accuracy of herceptin by fragmentation type and protease (Figure S3); herceptin peptide length distribution by digestion time (Figure S4); isoleucine/leucine assignment at heavy-chain position 51 of anti-FLAG-M2 (Figure S5); electrostatic surface potential of the anti-FLAG-M2 paratope (Figure S6); and Western blot validation of synthetic recombinant anti-FLAG-M2 (Figure S7) The raw LC-MS/MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD023419. The coordinates and reflection file with phases for the remodeled crystal structure of the anti-FLAG-M2 Fab have been deposited in the Protein Data Bank under accession code 7BG1 (PDF)

AUTHOR INFORMATION

Corresponding Author

Joost Snijder – Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands;
orcid.org/0000-0002-9310-8226; Email: j.snijder@uu.nl

Authors

Weiwei Peng – Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands

Matti F. Pronker – Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jproteome.1c00169>

Author Contributions

§W.P. and M.F.P. contributed equally. W.P. and J.S. conceived of the project. W.P. carried out the MS experiments. W.P. and J.S. analyzed the MS data. M.F.P. remodeled the crystal structure. M.F.P. cloned and produced the synthetic recombinant antibody and carried out Western blotting. J.S. supervised the project. J.S. wrote the first draft, and all authors contributed to preparing the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Herceptin was a kind gift from Roche (Penzberg, Germany). The authors would like to acknowledge support by Protein Metrics Inc. through access to Supernovo software and helpful discussion on *de novo* antibody sequencing. The authors would like to thank everyone in the Biomolecular Mass Spectrometry and Proteomics group at Utrecht University for support and helpful discussions. This research was funded by the Dutch Research Council NWO Gravitation 2013 BOO, Institute for Chemical Immunology (ICI; 024.002.009).

REFERENCES

- (1) Tonegawa, S. Somatic generation of antibody diversity. *Nature* **1983**, *302*, 575–581.
- (2) Watson, C. T.; Glanville, J.; Marasco, W. A. The individual and population genetics of antibody immunity. *Trends Immunol.* **2017**, *38*, 459–470.
- (3) Carter, P. J.; Lazar, G. A. Next generation antibody drugs: pursuit of the 'high-hanging fruit'. *Nat. Rev. Drug Discovery* **2018**, *17*, 197.
- (4) Grilo, A. L.; Mantalaris, A. The increasingly human and profitable monoclonal antibody market. *Trends Biotechnol.* **2019**, *37*, 9–16.
- (5) Baker, M. Blame it on the antibodies. *Nature* **2015**, *521*, 274.
- (6) Uhlen, M.; Bandrowski, A.; Carr, S.; Edwards, A.; Ellenberg, J.; Lundberg, E.; Rimm, D. L.; Rodriguez, H.; Hiltke, T.; Snyder, M.; et al. A proposal for validation of antibodies. *Nat. Methods* **2016**, *13*, 823–827.
- (7) Fischer, N. *Sequencing Antibody Repertoires: The Next Generation mAbs*; Taylor & Francis, 2011; pp 17–20.
- (8) 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*, 158–168.
- (9) Robinson, W. H. Sequencing the functional antibody repertoire—diagnostic and therapeutic discovery. *Nat. Rev. Rheumatol.* **2015**, *11*, 171.
- (10) Boutz, D. R.; Horton, A. P.; Wine, Y.; Lavinder, J. J.; Georgiou, G.; Marcotte, E. M. Proteomic identification of monoclonal antibodies from serum. *Anal. Chem.* **2014**, *86*, 4758–4766.
- (11) Castellana, N. E.; McCutcheon, K.; Pham, V. C.; Harden, K.; Nguyen, A.; Young, J.; Adams, C.; Schroeder, K.; Arnott, D.; Bafna, V.; et al. Resurrection of a clinical antibody: Template proteogenomic *de novo* proteomic sequencing and reverse engineering of an anti-lymphotoxin- α antibody. *Proteomics* **2011**, *11*, 395–405.
- (12) Chen, J.; Zheng, Q.; Hammers, C. M.; Ellebrecht, C. T.; Mukherjee, E. M.; Tang, H.-Y.; Lin, C.; Yuan, H.; Pan, M.; Langenhan, 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.* **2017**, *18*, 237–247.
- (13) 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.; et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat. Biotechnol.* **2012**, *30*, 447–452.
- (14) 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*, 45–54.
- (15) Lee, J.; Boutz, D. R.; Chromikova, V.; Joyce, M. G.; Vollmers, C.; Leung, K.; Horton, A. P.; DeKosky, B. J.; Lee, C.-H.; Lavinder, J. J.; et al. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat. Med.* **2016**, *22*, 1456–1464.
- (16) Lee, J.; Paparoditis, P.; Horton, A. P.; Frühwirth, A.; McDaniel, J. R.; Jung, J.; Boutz, D. R.; Hussein, D. A.; Tanno, Y.; Pappas, L.; et al. Persistent antibody clonotypes dominate the serum response to influenza over multiple years and repeated vaccinations. *Cell Host Microbe* **2019**, *25*, 367–376.
- (17) Lindesmith, L. C.; McDaniel, J. R.; Changela, A.; Verardi, R.; Kerr, S. A.; Costantini, V.; Brewer-Jensen, P. D.; Mallory, M. L.; Voss, W. N.; Boutz, D. R.; et al. Sera antibody repertoire analyses reveal mechanisms of broad and pandemic strain neutralizing responses after human norovirus vaccination. *Immunity* **2019**, *50*, 1530–1541.
- (18) Bandeira, N.; Pham, V.; Pevzner, P.; Arnott, D.; Lill, J. R. Automated *de novo* protein sequencing of monoclonal antibodies. *Nat. Biotechnol.* **2008**, *26*, 1336–1338.
- (19) 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*; Taylor & Francis, 2016; pp 501–512.
- (20) Savidor, A.; Barzilay, R.; Elinger, D.; Yarden, Y.; Lindzen, M.; Gabashvili, A.; Tal, O. A.; Levin, Y. Database-Independent Protein Sequencing (DiPS) Enables Full-Length *de Novo* Protein and Antibody Sequence Determination. *Mol. Cell. Proteomics* **2017**, *16*, 1151–1161.
- (21) 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*, 803–810.
- (22) 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.
- (23) 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*, No. 31730.
- (24) Brizzard, B. L.; Chubet, R. G.; Vizard, D. Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase

using a novel monoclonal antibody and peptide elution. *Biotechniques* **1994**, *16*, 730–735.

(25) Sigma-Aldrich anti-FLAG-M2 F1804 Product Page, 2021. <https://www.sigmaaldrich.com/catalog/product/sigma/f1804?lang=en&ion=NL>.

(26) Ehrenmann, F.; Kaas, Q.; Lefranc, M.-P. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res.* **2010**, *38*, D301–D307.

(27) Ehrenmann, F.; Lefranc, M.-P. IMGT/DomainGapAlign: IMGT standardized analysis of amino acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF). *Cold Spring Harbor Protoc.* **2011**, *2011*, 737–749.

(28) Roosild, T. P.; Castronovo, S.; Choe, S. Structure of anti-FLAG M2 Fab domain and its use in the stabilization of engineered membrane proteins. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2006**, *62*, 835–839.

(29) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

(30) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 352–367.

(31) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12–21.

(32) Fuglsang, A. Codon optimizer: a freeware tool for codon optimization. *Protein Expression Purif.* **2003**, *31*, 247–249.

(33) García-Nafria, J.; Watson, J. F.; Greger, I. H. IVA cloning: a single-tube universal cloning system exploiting bacterial in vivo assembly. *Sci. Rep.* **2016**, *6*, No. 27459.

(34) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508–548.

(35) Diedrich, J. K.; Pinto, A. F.; Yates, J. R., III Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1690–1699.

(36) Frese, C. K.; Altelaar, A. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012**, *84*, 9668–9673.

(37) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. M.; Mechtler, K.; Heck, A. J.; Mohammed, S. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **2013**, *12*, 1520–1525.

(38) Carter, P.; Presta, L.; Gorman, C. M.; Ridgway, J.; Henner, D.; Wong, W.; Rowland, A. M.; Kotts, C.; Carver, M. E.; Shepard, H. M. Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4285–4289.

(39) Slamon, D. J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **2001**, *344*, 783–792.

(40) Einhauer, A.; Jungbauer, A. The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **2001**, *49*, 455–465.

(41) Entzinger, K. C.; Hyun, J.-m.; Pantazes, R. J.; Patterson-Orazem, A. C.; Qerqez, A. N.; Frye, Z. P.; Hughes, R. A.; Ellington, A. D.; Lieberman, R. L.; Maranas, C. D.; et al. De novo design of antibody complementarity determining regions binding a FLAG tetrapeptide. *Sci. Rep.* **2017**, *7*, No. 10295.

(42) Ikeda, K.; Koga, T.; Sasaki, F.; Ueno, A.; Saeki, K.; Okuno, T.; Yokomizo, T. Generation and characterization of a human-mouse

chimeric high-affinity antibody that detects the DYKDDDDK FLAG peptide. *Biochem. Biophys. Res. Commun.* **2017**, *486*, 1077–1082.

(43) Lima, W. C.; Gasteiger, E.; Marcatili, P.; Duek, P.; Bairoch, A.; Cosson, P. The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Res.* **2020**, *48*, D261–D264.

(44) Bondt, A.; Hoek, M.; Tamara, S.; de Graaf, B.; Peng, W.; Schulte, D.; den Boer, M. A.; Greisch, J.-F.; Varkila, M. R.; Snijder, J.; et al. Human Plasma IgG1 Repertoires are Simple, Unique, and Dynamic. *SSRN Electron. J.* **2020**, DOI: 10.2139/ssrn.3749694.