

1 **Mass spectrometry-based sequencing of the anti-FLAG-M2 antibody using multiple**
2 **proteases and a dual fragmentation scheme**

3

4 **Authors:**

5 Weiwei Peng^{1#}, Matti F. Pronker^{1#}, Joost Snijder^{1*}

6

7 #equal contribution

8 *corresponding author: j.snijder@uu.nl

9

10 **Affiliation:**

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

14

15 **Keywords:**

16 mass spectrometry, antibody, de novo sequencing, EThcD, stepped HCD, Herceptin, FLAG tag,
17 anti-FLAG-M2.

18 **Abstract:**

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

31 Introduction

32 Antibodies can bind a great molecular diversity of antigens, owing to the high degree of sequence
33 diversity that is available through somatic recombination, hypermutation, and heavy-light chain
34 pairings¹⁻². Sequence information on antibodies therefore is crucial to understanding the
35 structural basis of antigen binding, how somatic hypermutation governs affinity maturation, and
36 an overall understanding of the adaptive immune response in health and disease, by mapping
37 out the antibody repertoire. Moreover, antibodies have become invaluable research tools in the
38 life sciences and ever more widely developed as therapeutic agents³⁻⁴. In this context, sequence
39 information is crucial for the use, production and validation of these important research tools and
40 biopharmaceutical agents⁵⁻⁶.

41 Antibody sequences are typically obtained through cloning and sequencing of the coding mRNAs
42 of the paired heavy and light chains⁷⁻⁹. The sequencing workflows thereby rely on isolation of the
43 antibody-producing cells from peripheral blood monocytes, or spleen and bone marrow tissues.
44 These antibody-producing cells are not always readily available however, and cloning/sequencing
45 of the paired heavy and light chains is a non-trivial task with a limited success rate⁷⁻⁹. Moreover,
46 antibodies are secreted in bodily fluids and mucus. Antibodies are thereby in large part
47 functionally disconnected from their producing B-cell, which raises questions on how the secreted
48 antibody pool relates quantitatively to the underlying B-cell population and whether there are
49 potential sampling biases in current antibody sequencing strategies.

50 Direct mass spectrometry (MS)-based sequencing of the secreted antibody products is a useful
51 complementary tool that can address some of the challenges faced by conventional sequencing
52 strategies relying on cloning/sequencing of the coding mRNAs¹⁰⁻¹⁷. MS-based methods do not
53 rely on the availability of the antibody-producing cells, but rather target the polypeptide products
54 directly, offering the prospect of a next generation of serology, in which secreted antibody
55 sequences might be obtained from any bodily fluid. Whereas MS-based *de novo* sequencing still
56 has a long way to go towards this goal, owing to limitations in sample requirements, sequencing
57 accuracy, read length and sequence assembly, MS has been successfully used to profile the
58 antibody repertoire and obtain (partial) antibody sequences beyond those available from
59 conventional sequencing strategies based on cloning/sequencing of the coding mRNAs¹⁰⁻¹⁷.

60 Most MS-based strategies for antibody sequencing rely on a proteomics-type bottom-up LC-
61 MS/MS workflow, in which the antibody product is digested into smaller peptides for MS analysis
62^{14, 18-23}. Available germline antibody sequences are then often used either as a template to guide

63 assembly of *de novo* peptide reads (such as in PEAKS Ab)²³, or used as a starting point to
64 iteratively identify somatic mutations to arrive at the mature antibody sequence (such as in
65 Supernovo)²¹. To maximize sequence coverage and aid read assembly, these MS-based
66 workflows typically use a combination of complementary proteases and aspecific digestion to
67 generate overlapping peptides. The most straightforward application of these MS-based
68 sequencing workflows is the successful sequencing of monoclonal antibodies from (lost)
69 hybridoma cell lines, but it also forms the basis of more advanced and challenging applications to
70 characterize polyclonal antibody mixtures and profile the full antibody repertoire from serum.

71 Here we describe an efficient protocol for MS-based sequencing of monoclonal antibodies. The
72 protocol requires approximately 200 picomol of the antibody product and sample preparation can
73 be completed within one working day. We selected a panel of 9 proteases with complementary
74 specificities, which are active in the same buffer conditions for parallel digestion of the antibodies.
75 We developed a dual fragmentation strategy for MS/MS analysis of the resulting peptides to yield
76 rich sequence information from the fragmentation spectra of the peptides. The protocol yields full
77 and deep sequence coverage of the variable domains of both heavy and light chains as
78 demonstrated on the monoclonal antibody Herceptin. As a test case, we used our protocol to
79 sequence the widely used anti-FLAG-M2 mouse monoclonal antibody, for which no sequence
80 was publicly available despite its described use in 5000+ peer-reviewed publications²⁴⁻²⁵. The
81 protocol achieved full sequence coverage of the variable domains of both heavy and light chains,
82 including all complementarity determining regions (CDRs). The obtained sequence was
83 successfully validated by remodeling the published crystal structure of the anti-FLAG-M2 Fab and
84 demonstrating binding of the synthetic recombinant antibody following the experimental sequence
85 to a FLAG-tagged protein in Western blot analysis. The protocol developed here thus offers robust
86 and reliable sequencing of monoclonal antibodies with prospective applications for sequencing
87 secreted antibodies from bodily fluids.

88

89 **Results**

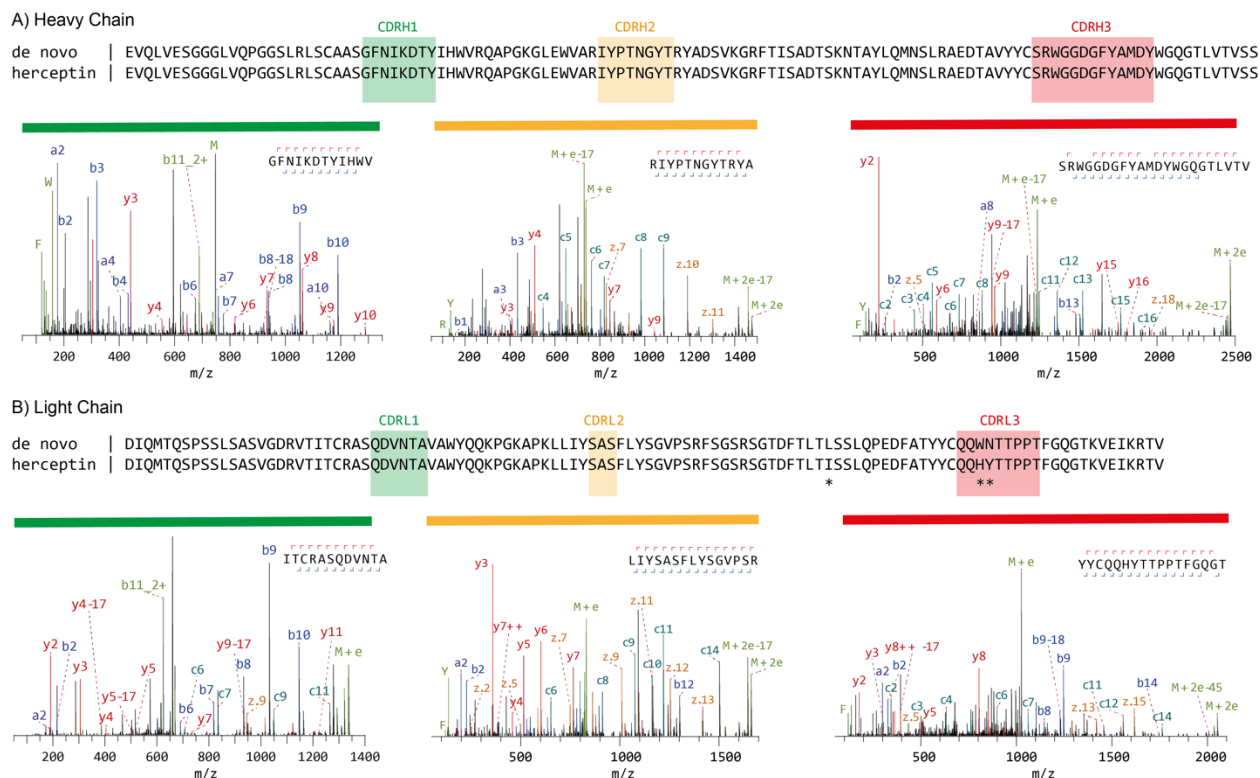
90 We used an in-solution digestion protocol, with sodium-deoxycholate as the denaturing agent, to
91 generate peptides from the antibodies for LC-MS/MS analysis. Following heat denaturation and
92 disulfide bond reduction, we used iodoacetic acid as the alkylating agent to cap free cysteines.
93 Note that conventional alkylating agents like iodo-/chloroacetamide generate +57 Da mass
94 differences on cysteines and primary amines, which may lead to spurious assignments as glycine

95 residues in *de novo* sequencing. The +58 Da mass differences generated by alkylation with
96 iodoacetic acid circumvents this potential pitfall.

97 We chose a panel of 9 proteases with activity at pH 7.5-8.5, so that the denatured, reduced and
98 alkylated antibodies could be easily split for parallel digestion under the same buffer conditions.
99 These proteases (with indicated cleavage specificities) included: trypsin (C-terminal of R/K),
100 chymotrypsin (C-terminal of F/Y/W/M/L), α -lytic protease (C-terminal of T/A/S/V), elastase
101 (unspecific), thermolysin (unspecific), lysN (N-terminal of K), lysC (C-terminal of K), aspN (N-
102 terminal of D/E), and gluC (C-terminal of D/E). Correct placement or assembly of peptide reads
103 is a common challenge in *de novo* sequencing, which can be facilitated by sufficient overlap
104 between the peptide reads. This favors the occurrence of missed cleavages and longer reads, so
105 we opted to perform a brief 4-hour digestion. Following digestion, SDC is removed by precipitation
106 and the peptide supernatant is desalted, ready for LC-MS/MS analysis. The resulting raw data
107 was used for automated *de novo* sequencing with the Supernovo software package.

108 As peptide fragmentation is dependent on many factors like length, charge state, composition and
109 sequence ²⁶, we needed a versatile fragmentation strategy to accommodate the diversity of
110 antibody-derived peptides generated by the 9 proteases. We opted for a dual fragmentation
111 scheme that applies both stepped high-energy collision dissociation (stepped HCD) and electron
112 transfer high-energy collision dissociation (ETHcD) on all peptide precursors ²⁷⁻²⁹. The stepped
113 HCD fragmentation includes three collision energies to cover multiple dissociation regimes and
114 the ETHcD fragmentation works especially well for higher charge states, also adding
115 complementary *c/z* ions for maximum sequence coverage.

116 We used the monoclonal antibody Herceptin (also known as Trastuzumab) as a benchmark to
117 test our protocol ³⁰⁻³¹. From the total dataset of 9 proteases, we collected 4408 peptide reads
118 (defined as peptides with score ≥ 500 , see methods for details), 2866 of which with superior
119 stepped HCD fragmentation, and 1722 with superior ETHcD fragmentation (see Table S1).
120 Sequence coverage was 100% in both heavy and light chains across the variable and constant
121 domains (see Figures S1 and S2). The median depth of coverage was 148 overall and slightly
122 higher in the light chain (see Table S1 and Figure S1-2). The median depth of coverage in the
123 CDRs of both chains ranged from 42 to 210.



124

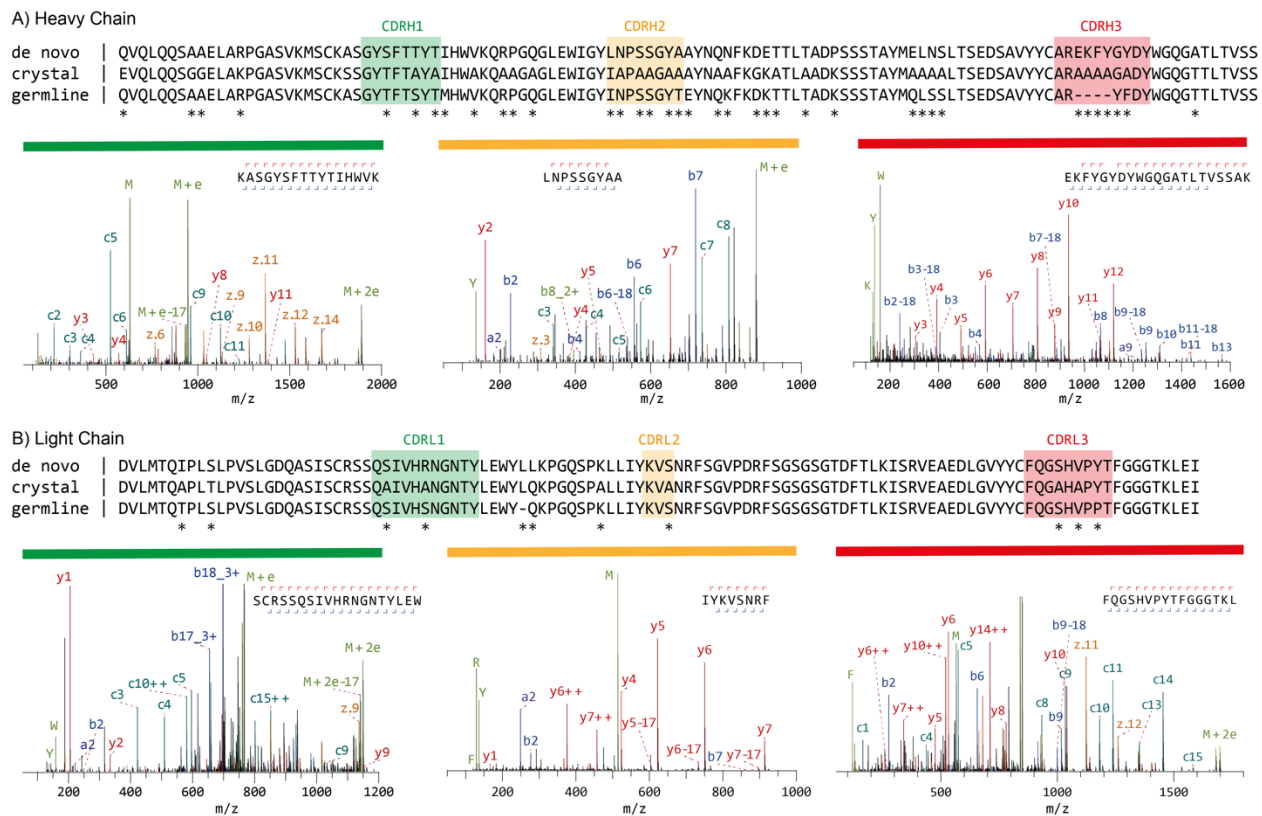
125 **Figure 1.** mass spectrometry-based *de novo* sequencing of the monoclonal antibody Herceptin. The
 126 variable regions of the Heavy (A) and Light Chains (B) are shown. The MS-based sequence is shown
 127 alongside the known Herceptin CDR sequence, with differences highlighted by asterisks (*). Exemplary MS/MS
 128 spectra supporting the assigned sequences of the Heavy and Light Chain CDRs are shown below the
 129 alignments. Peptide sequence and fragment coverage are indicated on top of the spectra, with b/c ions
 130 indicated in blue and y/z ions in red. The same coloring is used to annotate peaks in the spectra, with
 131 additional peaks such as intact/charge reduced precursors, neutral losses and immonium ions indicated in
 132 green. Note that to prevent overlapping peak labels, only a subset of successfully matched peaks is
 133 annotated.

134

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

143 supported by the fragmentation spectra, in contrast to the correct H91/Y92 assignment (see c6/c7
 144 in fragmentation spectra, Figure 1). Overall, the protocol yielded highly accurate sequences at a
 145 combined 230/233 positions of the variable domains in Herceptin.

146

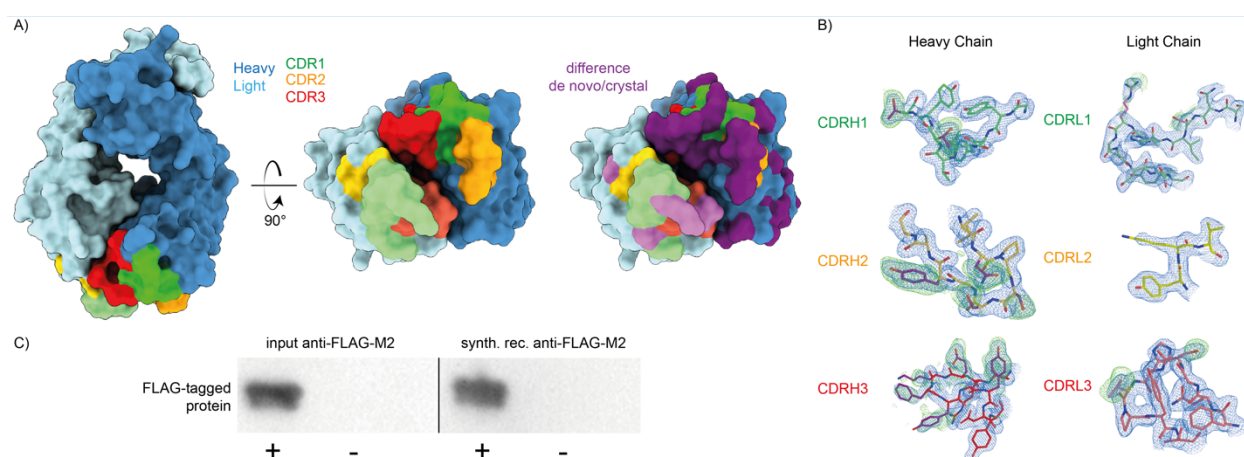


147
 148 **Figure 2.** Mass spectrometry based *de novo* sequence of the mouse monoclonal anti-FLAG-M2 antibody.
 149 The variable regions of the Heavy (A) and Light Chains (B) are shown. The MS-based sequence is shown
 150 alongside the previously published sequenced in the crystal structure of the Fab (PDB ID: 2G60), and
 151 germline sequence (IMGT-DomainGapAlign; IGHV1-04/IGHJ2; IGKV1-117/IGKJ1). Differential residues
 152 are highlighted by asterisks (*). Exemplary MS/MS spectra in support of the assigned sequences are shown
 153 below the alignments. Peptide sequence and fragment coverage are indicated on top of the spectra, with
 154 b/c ions indicated in blue, y/z ions in red. The same coloring is used to annotate peaks in the spectra, with
 155 additional peaks such as intact/charge reduced precursors, neutral losses and immonium ions indicated in
 156 green. Note that to prevent overlapping peak labels, only a subset of successfully matched peaks is
 157 annotated.

158

159 We next applied our sequencing protocol to the mouse monoclonal anti-FLAG-M2 antibody as a
160 test case ²⁴. Despite the widespread use of anti-FLAG-M2 to detect and purify FLAG-tagged
161 proteins ³², the only publicly available sequences can be found in the crystal structure of the Fab
162 ³³. The modelled sequence of the original crystal structure had to be inferred from germline
163 sequences that could match the experimental electron density and also includes many
164 placeholder Alanines at positions that could not be straightforwardly interpreted. The full anti-
165 FLAG-M2 dataset from the 9 proteases included 3371 peptide reads (with scores ≥ 500); 1983
166 with superior stepped HCD fragmentation spectra, and 1388 with superior ETHcD spectra. We
167 achieved full sequence coverage of the variable regions of both heavy and light chains, with a
168 median depth of coverage in the CDRs ranging from 32 to 192 (see Table S1). As for Herceptin,
169 the depth of coverage was better in the light chain compared to the heavy chain (see Figure S1-
170 S2). The full MS-based anti-FLAG-M2 sequences can be found in FASTA format in the
171 supplementary information.

172



173

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

183

184 The MS-based sequences of anti-FLAG-M2 are shown alongside the crystal structure sequences
185 and the inferred germline precursors with exemplary MS/MS spectra for the CDRs in Figure 2.
186 The experimentally determined sequence reveals that anti-FLAG-M2 is a mouse IgG1, with an
187 IGHV1-04/IGHJ2 heavy chain and IGKV1-117/IGKJ1 kappa light chain. The experimentally
188 determined sequence differs at 34 and 9 positions in the heavy and light chain of the Fab crystal
189 structure, respectively. To validate the experimentally determined sequences, we remodeled the
190 crystal structure using the MS-based heavy and light chains, resulting in much improved model
191 statistics (see Figure 3 and Table S2). The experimental electron densities show excellent support
192 of the MS-based sequence (as shown for the CDRs in Figure 3B). A notable exception is L51 in
193 CDRH2 of the heavy chain. The MS-based sequence was assigned as Leucine, but the
194 experimental electron density supports assignment of the isomer Isoleucine instead (see Figure
195 S3). In contrast to the original model our new MS-based model reveals a predominantly positively
196 charged paratope (see Figure S4), which potentially complements the -3 net charge of the FLAG
197 tag epitope (DYKDDDDK) to mediate binding. The experimentally determined anti-FLAG-M2
198 sequence, with the L51I correction, was further validated by testing binding of the synthetic
199 recombinant antibody to a purified FLAG-tagged protein in Western blot analysis (see Figure 3C
200 and S5). The synthetic recombinant antibody showed equivalent binding compared to the original
201 antibody sample used for sequencing, confirming that the experimentally determined sequence
202 is reliable to obtain the recombinant antibody product with the desired functional profile.

203

204 Discussion

205 There are four other monoclonal antibody sequences against the FLAG tag publicly available
206 through the ABCD (AntiBodies Chemically Defined) database³⁴⁻³⁶. Comparison of the CDRs of
207 anti-FLAG-M2 with these additional four monoclonal antibodies reveals a few common motifs that
208 may determine FLAG-tag binding specificity (see Table S3). In the heavy chain, the only common
209 motif between all five monoclonals is that the first three residues of CDRH1 follow a GXS
210 sequence. In addition, the last three residues of CDRH3 of anti-FLAG-M2 are YDY, similar to
211 MDY in H28, and YDF in EEh13.6 (and EEh14.3 also ends CDRH3 with an aromatic F residue).
212 In contrast to the heavy chain, the CDRs of the light chain are almost completely conserved in
213 4/5 monoclonals with only minimal differences compared to germline. The anti-FLAG-M2 and H28
214 monoclonals were specifically raised in mice against the FLAG-tag epitope^{24, 35}, whereas the
215 computationally designed EEh13.6 and EEh14.3 monoclonals contain the same light chain from
216 an EE-dipeptide tag directed antibody³⁴. This suggests that the IGKV1-117/IGKJ1 light chain may

217 be a common determinant of binding to a small negatively charged peptide epitope like the FLAG-
218 tag and is readily available as a hardcoded germline sequence in the mouse antibody repertoire.

219 The availability of the anti-FLAG-M2 sequences may contribute to the wider use of this important
220 research tool, as well as the development and engineering of better FLAG-tag directed antibodies.
221 This example illustrates that our MS-based sequencing protocol yields robust and reliable
222 monoclonal antibody sequences. The protocol described here also formed the basis of a recent
223 application where we sequenced an antibody directly from patient-derived serum, using a
224 combination with top-down fragmentation of the isolated Fab fragment³⁷. The dual fragmentation
225 strategy yields high-quality spectra suitable for *de novo* sequencing and may further contribute to
226 the exciting prospect of a new era of serology in which antibody sequences can be directly
227 obtained from bodily fluids.

228

229

230 **Methods**

231 *Sample preparation*

232 Anti-Flag M2 antibody was purchased from Sigma (catalogue number F1804). Herceptin was
233 provided by Roche (Penzberg, Germany). 27 µg of each sample was denatured in 2% sodium
234 deoxycholate (SDC), 200 mM Tris-HCl, 10 mM tris(2-carboxyethyl)phosphine (TCEP), pH 8.0 at
235 95°C for 10 min, followed with 30 min incubation at 37°C for reduction. Sample was then alkylated
236 by adding iodoacetic acid to a final concentration of 40 mM and incubated in the dark at room
237 temperature for 45 min. 3 µg Sample was then digested by one of the following proteases: trypsin,
238 chymotrypsin, lysN, lysC, gluC, aspN, aLP, thermolysin and elastase in a 1:50 ratio (w:w) in a
239 total volume of 100 µL of 50 mM ammonium bicarbonate at 37°C for 4 h. After digestion, SDC
240 was removed by adding 2 µL formic acid (FA) and centrifugation at 14000 g for 20 min. Following
241 centrifugation, the supernatant containing the peptides was collected for desalting on a 30 µm
242 Oasis HLB 96-well plate (Waters). The Oasis HLB sorbent was activated with 100% acetonitrile
243 and subsequently equilibrated with 10% formic acid in water. Next, peptides were bound to the
244 sorbent, washed twice with 10% formic acid in water and eluted with 100 µL of 50%
245 acetonitrile/5% formic acid in water (v/v). The eluted peptides were vacuum-dried and
246 reconstituted in 100 µL 2% FA.

247

248 *Mass Spectrometry*

249 The digested peptides (single injection of 0.2 ug) were separated by online reversed phase
250 chromatography on an Agilent 1290 UHPLC (column packed with Poroshell 120 EC C18;
251 dimensions 50 cm x 75 μ m, 2.7 μ m, Agilent Technologies) coupled to a Thermo Scientific Orbitrap
252 Fusion mass spectrometer. Samples were eluted over a 90 min gradient from 0% to 35%
253 acetonitrile at a flow rate of 0.3 μ L/min. Peptides were analyzed with a resolution setting of 60000
254 in MS1. MS1 scans were obtained with standard AGC target, maximum injection time of 50 ms,
255 and scan range 350-2000. The precursors were selected with a 3 m/z window and fragmented by
256 stepped HCD as well as EThcD. The stepped HCD fragmentation included steps of 25%, 35%
257 and 50% NCE. EThcD fragmentation was performed with calibrated charge-dependent ETD
258 parameters and 27% NCE supplemental activation. For both fragmentation types, ms2 scan were
259 acquired at 30000 resolution, 800% Normalized AGC target, 250 ms maximum injection time,
260 scan range 120-3500.

261

262 *MS Data Analysis*

263 Automated *de novo* sequencing was performed with Supernovo (version 3.10, Protein Metrics
264 Inc.). Custom parameters were used as follows: non-specific digestion; precursor and product
265 mass tolerance was set to 12 ppm and 0.02 Da respectively; carboxymethylation (+58.005479)
266 on cysteine was set as fixed modification; oxidation on methionine and tryptophan was set as
267 variable common 1 modification; carboxymethylation on the N-terminus, pyroglutamic acid
268 conversion of glutamine and glutamic acid on the N-terminus, deamidation on
269 asparagine/glutamine were set as variable rare 1 modifications. Peptides were filtered for score
270 ≥ 500 for the final evaluation of spectrum quality and (depth of) coverage. Supernovo generates
271 peptide groups for redundant MS/MS spectra, including also when stepped HCD and EThcD
272 fragmentation on the same precursor both generate good peptide-spectrum matches. In these
273 cases only the best-matched spectrum is counted as representative for that group. This criterium
274 was used in counting the number of peptide reads reported in Table S1. Germline sequences and
275 CDR boundaries were inferred using IMGT/DomainGapAlign³⁸⁻³⁹.

276

277

278

279 *Revision of the anti-FLAG-M2 Fab crystal structure model*

280 As a starting point for model building, the reflection file and coordinates of the published anti-
281 FLAG-M2 Fab crystal structure were used (PDB ID: 2G60)³³. Care was taken to use the original
282 R_{free} labels of the deposited reflection file for refinement, so as not to introduce extra model bias.
283 Differential residues between this structure and our mass spectrometry-derived anti-FLAG
284 sequence were manually mutated and fitted in the density using Coot⁴⁰. Many spurious water
285 molecules that caused severe steric clashes in the original model were also manually removed in
286 Coot. Densities for two sulfate and one chloride ion were identified and built into the model. The
287 original crystallization solution contained 0.1 M ammonium sulfate. Iterative cycles of model
288 geometry optimization in real space in Coot and reciprocal space refinement by Phenix were used
289 to generate the final model, which was validated with Molprobit⁴¹⁻⁴².

290

291 *Cloning and expression of synthetic recombinant anti-FLAG-M2*

292 To recombinantly express full-length anti-FLAG-M2, the proteomic sequences of both the light
293 and heavy chains were reverse-translated and codon optimized for expression in human cells
294 using the Integrated DNA Technologies (IDT) web tool (<http://www.idtdna.com/CodonOpt>)⁴³. For
295 the linker and Fc region of the heavy chain, the standard mouse Ig gamma-1 (IGHG1) amino acid
296 sequence (Uniprot P01868.1) was used. An N-terminal secretion signal peptide derived from
297 human IgG light chain (MEAPAQLLFLLLLWLPDTTG) was added to the N-termini of both heavy
298 and light chains. BamHI and NotI restriction sites were added to the 5' and 3' ends of the coding
299 regions, respectively. Only for the light chain, a double stop codon was introduced at the 3' site
300 before the NotI restriction site. The coding regions were subcloned using BamHI and NotI
301 restriction-ligation into a pRK5 expression vector with a C-terminal octahistidine tag between the
302 NotI site and a double stop codon 3' of the insert, so that only the heavy chain has a C-terminal
303 AAAHHHHHHHH sequence for Nickel-affinity purification (the triple alanine resulting from the NotI
304 site). The L51I correction in the heavy chain was introduced later (after observing it in the crystal
305 structure) by IVA cloning⁴⁴. Expression plasmids for the heavy and light chain were mixed in a
306 1:1 (w/w) ratio for transient transfection in HEK293 cells with polyethylenimine, following standard
307 procedures. Medium was collected 6 days after transfection and cells were spun down by
308 10 minutes of centrifugation at 1000 g. Antibody was directly purified from the supernatant using
309 Ni-sepharose excel resin (Cytiva Lifes Sciences), washing with 500 mM NaCl, 2 mM CaCl₂, 15

310 mM imidazole, 20 mM HEPES pH 7.8 and eluting with 500 mM NaCl, 2 mM CaCl₂, 200 mM
311 imidazole, 20 mM HEPES pH 7.8.

312

313 *Western blot validation of anti-FLAG-M2 binding*

314 To test binding of our recombinant anti-FLAG-M2 to the FLAG-tag epitope, compared to the
315 commercially available anti-FLAG-M2 (Sigma), we used both antibodies to probe Western blots
316 of a FLAG-tagged protein in parallel. Purified Rabies virus glycoprotein ectodomain (SAD B19
317 strain, UNIPROT residues 20-450) with or without a C-terminal FLAG-tag followed by a foldon
318 trimerization domain and an octahistidine tag was heated to 95 °C in XT sample buffer (Biorad)
319 for 5 minutes. Samples were run twice on a Criterion XT 4-12% polyacrylamide gel (Biorad) in
320 MES XT buffer (Biorad) before Western blot transfer to a nitrocellulose membrane in tris-glycine
321 buffer (Biorad) with 20% methanol. The membrane was blocked with 5% (w/v) dry non-fat milk in
322 phosphate-buffered saline (PBS) overnight at 4 °C. The membrane was cut in two (one half for
323 the commercial and one half for the recombinant anti-FLAG-M2) and each half was probed with
324 either commercial (Sigma) or recombinant anti-FLAG-M2 at 1 µg/mL in PBS for 45 minutes. After
325 washing three times with PBST (PBS with 0.1% v/v Tween20), polyclonal goat anti-mouse fused
326 to horseradish peroxidase (HRP) was used to detect binding of anti-FLAG-M2 to the FLAG-tagged
327 protein for both membranes. The membranes were washed three more times with PBST before
328 applying enhanced chemiluminescence (ECL; Pierce) reagent to image the blots in parallel.

329

330 **Data Availability**

331 The raw LC-MS/MS data have been deposited to the ProteomeXchange Consortium via the
332 PRIDE partner repository with the dataset identifier PXD023419. The coordinates and reflection
333 file with phases for the remodeled crystal structure of the anti-FLAG-M2 Fab have been deposited
334 in the Protein Data Bank under accession code 7BG1.

335

336 **Acknowledgements**

337 Herceptin was a kind gift from Roche (Penzberg, Germany). We would like to acknowledge
338 support by Protein Metrics Inc. through access to Supernovo software and helpful discussion on
339 *de novo* antibody sequencing. We would like to thank everyone in the Biomolecular Mass

340 Spectrometry and Proteomics group at Utrecht University for support and helpful discussions.
341 This research was funded by the Dutch Research Council NWO Gravitation 2013 BOO, Institute
342 for Chemical Immunology (ICI; 024.002.009).

343

344 **Author Contributions**

345 WP and JS conceived of the project. WP carried out the MS experiments. WP and JS analyzed
346 the MS data. MFP remodeled the crystal structure. MFP cloned and produced the synthetic
347 recombinant antibody and carried out Western blotting. JS supervised the project. JS wrote the
348 first draft and all authors contributed to preparing the final version of the manuscript.

349

350 **Competing Interests**

351 The authors declare no competing interests

352

353 **References**

- 354 1. Tonegawa, S., Somatic generation of antibody diversity. *Nature* **1983**, *302* (5909), 575-
355 581.
- 356 2. Watson, C. T.; Glanville, J.; Marasco, W. A., The individual and population genetics of
357 antibody immunity. *Trends in immunology* **2017**, *38* (7), 459-470.
- 358 3. Carter, P. J.; Lazar, G. A., Next generation antibody drugs: pursuit of the 'high-hanging
359 fruit'. *Nature Reviews Drug Discovery* **2018**, *17* (3), 197.
- 360 4. Grilo, A. L.; Mantalaris, A., The increasingly human and profitable monoclonal antibody
361 market. *Trends in biotechnology* **2019**, *37* (1), 9-16.
- 362 5. Baker, M., Blame it on the antibodies. *Nature* **2015**, *521* (7552), 274.
- 363 6. Uhlen, M.; Bandrowski, A.; Carr, S.; Edwards, A.; Ellenberg, J.; Lundberg, E.; Rimm, D.
364 L.; Rodriguez, H.; Hiltke, T.; Snyder, M., A proposal for validation of antibodies. *Nature methods*
365 **2016**, *13* (10), 823-827.
- 366 7. Fischer, N. In *Sequencing antibody repertoires: the next generation*, MAbs, Taylor &
367 Francis: 2011; pp 17-20.
- 368 8. Georgiou, G.; Ippolito, G. C.; Beausang, J.; Busse, C. E.; Wardemann, H.; Quake, S. R.,
369 The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature*
370 *biotechnology* **2014**, *32* (2), 158-168.

- 371 9. Robinson, W. H., Sequencing the functional antibody repertoire—diagnostic and
372 therapeutic discovery. *Nature Reviews Rheumatology* **2015**, *11* (3), 171.
- 373 10. Boutz, D. R.; Horton, A. P.; Wine, Y.; Lavinder, J. J.; Georgiou, G.; Marcotte, E. M.,
374 Proteomic identification of monoclonal antibodies from serum. *Analytical chemistry* **2014**, *86*
375 (10), 4758-4766.
- 376 11. Castellana, N. E.; McCutcheon, K.; Pham, V. C.; Harden, K.; Nguyen, A.; Young, J.;
377 Adams, C.; Schroeder, K.; Arnott, D.; Bafna, V., Resurrection of a clinical antibody: Template
378 proteogenomic de novo proteomic sequencing and reverse engineering of an anti-lymphotoxin-
379 α antibody. *Proteomics* **2011**, *11* (3), 395-405.
- 380 12. Chen, J.; Zheng, Q.; Hammers, C. M.; Ellebrecht, C. T.; Mukherjee, E. M.; Tang, H.-Y.;
381 Lin, C.; Yuan, H.; Pan, M.; Langenhan, J., Proteomic analysis of pemphigus autoantibodies
382 indicates a larger, more diverse, and more dynamic repertoire than determined by B cell
383 genetics. *Cell reports* **2017**, *18* (1), 237-247.
- 384 13. Cheung, W. C.; Beausoleil, S. A.; Zhang, X.; Sato, S.; Schieferl, S. M.; Wieler, J. S.;
385 Beaudet, J. G.; Ramenani, R. K.; Popova, L.; Comb, M. J., A proteomics approach for the
386 identification and cloning of monoclonal antibodies from serum. *Nature biotechnology* **2012**, *30*
387 (5), 447-452.
- 388 14. Guthals, A.; Gan, Y.; Murray, L.; Chen, Y.; Stinson, J.; Nakamura, G.; Lill, J. R.;
389 Sandoval, W.; Bandeira, N., De novo MS/MS sequencing of native human antibodies. *Journal of*
390 *proteome research* **2017**, *16* (1), 45-54.
- 391 15. Lee, J.; Boutz, D. R.; Chromikova, V.; Joyce, M. G.; Vollmers, C.; Leung, K.; Horton, A.
392 P.; DeKosky, B. J.; Lee, C.-H.; Lavinder, J. J., Molecular-level analysis of the serum antibody
393 repertoire in young adults before and after seasonal influenza vaccination. *Nature medicine*
394 **2016**, *22* (12), 1456-1464.
- 395 16. Lee, J.; Paparoditis, P.; Horton, A. P.; Frühwirth, A.; McDaniel, J. R.; Jung, J.; Boutz, D.
396 R.; Hussein, D. A.; Tanno, Y.; Pappas, L., Persistent antibody clonotypes dominate the serum
397 response to influenza over multiple years and repeated vaccinations. *Cell host & microbe* **2019**,
398 *25* (3), 367-376. e5.
- 399 17. Lindesmith, L. C.; McDaniel, J. R.; Changela, A.; Verardi, R.; Kerr, S. A.; Costantini, V.;
400 Brewer-Jensen, P. D.; Mallory, M. L.; Voss, W. N.; Boutz, D. R., Sera antibody repertoire
401 analyses reveal mechanisms of broad and pandemic strain neutralizing responses after human
402 norovirus vaccination. *Immunity* **2019**, *50* (6), 1530-1541. e8.
- 403 18. Bandeira, N.; Pham, V.; Pevzner, P.; Arnott, D.; Lill, J. R., Automated de novo protein
404 sequencing of monoclonal antibodies. *Nature biotechnology* **2008**, *26* (12), 1336-1338.
- 405 19. Rickert, K. W.; Grinberg, L.; Woods, R. M.; Wilson, S.; Bowen, M. A.; Baca, M. In
406 *Combining phage display with de novo protein sequencing for reverse engineering of*
407 *monoclonal antibodies*, MABs, Taylor & Francis: 2016; pp 501-512.
- 408 20. Savidor, A.; Barzilay, R.; Elinger, D.; Yarden, Y.; Lindzen, M.; Gabashvili, A.; Tal, O. A.;
409 Levin, Y., Database-Independent Protein Sequencing (DiPS) Enables Full-Length de Novo
410 Protein and Antibody Sequence Determination. *Molecular & Cellular Proteomics* **2017**, *16* (6),
411 1151-1161.

- 412 21. Sen, K. I.; Tang, W. H.; Nayak, S.; Kil, Y. J.; Bern, M.; Ozoglu, B.; Ueberheide, B.; Davis,
413 D.; Becker, C., Automated antibody de novo sequencing and its utility in biopharmaceutical
414 discovery. *Journal of The American Society for Mass Spectrometry* **2017**, *28* (5), 803-810.
- 415 22. Sousa, E.; Olland, S.; Shih, H. H.; Marquette, K.; Martone, R.; Lu, Z.; Paulsen, J.; Gill,
416 D.; He, T., Primary sequence determination of a monoclonal antibody against α -synuclein using
417 a novel mass spectrometry-based approach. *International Journal of Mass Spectrometry* **2012**,
418 *312*, 61-69.
- 419 23. Tran, N. H.; Rahman, M. Z.; He, L.; Xin, L.; Shan, B.; Li, M., Complete de novo assembly
420 of monoclonal antibody sequences. *Scientific reports* **2016**, *6* (1), 1-10.
- 421 24. Brizzard, B. L.; Chubet, R. G.; Vizard, D., Immunoaffinity purification of FLAG epitope-
422 tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution.
423 *Biotechniques* **1994**, *16* (4), 730-735.
- 424 25. Sigma-Aldrich anti-FLAG-M2 F1804 product page.
425 <https://www.sigmaaldrich.com/catalog/product/sigma/f1804?lang=en®ion=NL> (accessed 05-
426 01-2021).
- 427 26. Paizs, B.; Suhai, S., Fragmentation pathways of protonated peptides. *Mass*
428 *spectrometry reviews* **2005**, *24* (4), 508-548.
- 429 27. Diedrich, J. K.; Pinto, A. F.; Yates III, J. R., Energy dependence of HCD on peptide
430 fragmentation: stepped collisional energy finds the sweet spot. *Journal of the American Society*
431 *for Mass Spectrometry* **2013**, *24* (11), 1690-1699.
- 432 28. Frese, C. K.; Altelaar, A. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A.
433 J.; Mohammed, S., Toward full peptide sequence coverage by dual fragmentation combining
434 electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Analytical*
435 *chemistry* **2012**, *84* (22), 9668-9673.
- 436 29. Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. M.; Mechtler, K.; Heck, A. J.; Mohammed,
437 S., Unambiguous phosphosite localization using electron-transfer/higher-energy collision
438 dissociation (ET_hcD). *Journal of proteome research* **2013**, *12* (3), 1520-1525.
- 439 30. Carter, P.; Presta, L.; Gorman, C. M.; Ridgway, J.; Henner, D.; Wong, W.; Rowland, A.
440 M.; Kotts, C.; Carver, M. E.; Shepard, H. M., Humanization of an anti-p185HER2 antibody for
441 human cancer therapy. *Proceedings of the National Academy of Sciences* **1992**, *89* (10), 4285-
442 4289.
- 443 31. Slamon, D. J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.;
444 Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M., Use of chemotherapy plus a monoclonal
445 antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England*
446 *journal of medicine* **2001**, *344* (11), 783-792.
- 447 32. Einhauer, A.; Jungbauer, A., The FLAG™ peptide, a versatile fusion tag for the
448 purification of recombinant proteins. *Journal of biochemical and biophysical methods* **2001**, *49*
449 (1-3), 455-465.

- 450 33. Roosild, T. P.; Castronovo, S.; Choe, S., Structure of anti-FLAG M2 Fab domain and its
451 use in the stabilization of engineered membrane proteins. *Acta Crystallographica Section F:*
452 *Structural Biology and Crystallization Communications* **2006**, 62 (9), 835-839.
- 453 34. Entzminger, K. C.; Hyun, J.-m.; Pantazes, R. J.; Patterson-Orazem, A. C.; Qerqez, A. N.;
454 Frye, Z. P.; Hughes, R. A.; Ellington, A. D.; Lieberman, R. L.; Maranas, C. D., De novo design of
455 antibody complementarity determining regions binding a FLAG tetra-peptide. *Scientific reports*
456 **2017**, 7 (1), 1-11.
- 457 35. Ikeda, K.; Koga, T.; Sasaki, F.; Ueno, A.; Saeki, K.; Okuno, T.; Yokomizo, T., Generation
458 and characterization of a human-mouse chimeric high-affinity antibody that detects the
459 DYKDDDDK FLAG peptide. *Biochemical and Biophysical Research Communications* **2017**, 486
460 (4), 1077-1082.
- 461 36. Lima, W. C.; Gasteiger, E.; Marcatili, P.; Duek, P.; Bairoch, A.; Cosson, P., The ABCD
462 database: a repository for chemically defined antibodies. *Nucleic acids research* **2020**, 48 (D1),
463 D261-D264.
- 464 37. Bondt, A.; Hoek, M.; Tamara, S.; de Graaf, B.; Peng, W.; Schulte, D.; den Boer, M. A.;
465 Greisch, J.-F.; Varkila, M. R.; Snijder, J., Human Plasma IgG1 Repertoires are Simple, Unique,
466 and Dynamic. *SSRN* **2020**.
- 467 38. Ehrenmann, F.; Kaas, Q.; Lefranc, M.-P., IMGT/3Dstructure-DB and
468 IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell
469 receptors, MHC, IgSF and MhcSF. *Nucleic acids research* **2010**, 38 (suppl_1), D301-D307.
- 470 39. Ehrenmann, F.; Lefranc, M.-P., IMGT/DomainGapAlign: IMGT standardized analysis of
471 amino acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF).
472 *Cold Spring Harbor Protocols* **2011**, 2011 (6), pdb. prot5636.
- 473 40. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta*
474 *Crystallographica Section D: Biological Crystallography* **2004**, 60 (12), 2126-2132.
- 475 41. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.;
476 Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards
477 automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica*
478 *Section D: Biological Crystallography* **2012**, 68 (4), 352-367.
- 479 42. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G.
480 J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation
481 for macromolecular crystallography. *Acta Crystallographica Section D: Biological*
482 *Crystallography* **2010**, 66 (1), 12-21.
- 483 43. Fuglsang, A., Codon optimizer: a freeware tool for codon optimization. *Protein*
484 *expression and purification* **2003**, 31 (2), 247-249.
- 485 44. García-Nafría, J.; Watson, J. F.; Greger, I. H., IVA cloning: a single-tube universal
486 cloning system exploiting bacterial in vivo assembly. *Scientific reports* **2016**, 6, 27459.

487

488