Complete *de novo* sequencing of a human immunoglobulin G using multiple enzyme digests and LC-ESI-QTOF-MS/MS analysis

Detailed method for antibody *de novo* sequencing

Abstract

The development of monoclonal antibody-based biotherapeutics requires the determination of the full protein sequence. The cloning of antibodies typically provides a reference sequence that can be relatively easily confirmed. Thus, the development of biosimilars often depends on the determination of the full originator sequence *de novo*. In many diagnostically or therapeutically relevant cases, the full DNA-sequence of an IgG of interest is not available. Here, we demonstrate complete *de novo* sequencing of a human monoclonal IgG by combining LC-ESI-QTOF-MS results from digests of a model monoclonal antibody with up to seven separate enzymes. LC-MS/MS results are combined and interpreted using PEAKS AB software.

Keywords: impact II QTOF, PEAKS AB software, mass spectrometry, de novo *sequencing, Biopharma, monoclonal antibodies, CDRs*

Introduction

Mass spectrometry *de novo* protein sequencing is an important tool for mAb development and can supplement database and in silico approaches, which rely on prior knowledge of the

Authors: Jinrui Gan[1], Tim Welsink[2], Detlev Suckau[3], Carsten Hopf[1]; [1]Center for Mass Spectrometry and Optical Spectroscopy (CeMOS), Mannheim Technical University, Mannheim, Germany; [2]InVivo BioTech Services GmbH, Hennigsdorf, Germany; [3]Bruker Daltonik GmbH, Bremen, Germany.

protein sequence. *De novo* mAb sequencing may also help when no cDNA for a given antibody expressed by a hybridoma cell line is available, the hybridoma cell line is lost or affected by genetic drift. Effective automated *de novo* protein sequencing of mAbs by LC-MS was first introduced in 2008 [1]. Nevertheless, analytical workflows and bioinformatics solutions required for *de novo* sequencing are still not widely established in laboratories around the globe.

A recombinant human IgG1 monoclonal antibody (AF165, InVivo Biotech) was used for this study. Brief description of its production: after transfection of suspension Chinese hamster ovary (CHO) cells, transfectants were selected and monoclonal antibody-producing cell lines isolated by a routine single cell cloning procedure, followed by an optimized fed-batch production with a high-producer clone. The antibody was purified from cell culture supernatant by a high purity protein A affinity chromatography method.

In the presented work, a detailed method for antibody *de novo* sequencing is described. It is based on LC-MS/MS analysis of multiple proteolytic enzyme digestion reactions to generate a panel of overlapping peptide sequences on a high resolution QTOF. Using dedicated antibody sequencing software, in a case study the sequence of a monoclonal antibody was determined, which was not known at the time of analysis.

Experimental

Prepare all solutions using ultrapure water using a water purification system, to attain a sensitivity of 18 MΩ-cm at 25°C) and analytical grade reagents (unless mentioned otherwise). Prepare and store all reagents at room temperature (unless indicated otherwise). Diligently follow all waste disposal regulations when disposing waste materials. We do not add sodium azide to reagents.

Table 1: List of materials used. Storage temperature is also provided.

Material	Order Information/Details
IgG Sample	AF165 antibody, 1.00 mg/mL, purified, in PBS pH 7.4, 2-8°C (InVivo BioTech Services GmbH)
PNGase F	PNGase F Glycosidase (N-zyme scientifics, PNGase F PRIME), -20°C
HCl	Hydrochloric acid (VWR #E447), RT
DTT	1,4-Dithiothreitol (PanReac AppliChem #A1101), 100 mM freshly prepared daily, RT
IAA	Iodoacetamide (Sigma #1149), 100 mM freshly prepared daily, RT
FA	Formic acid (VWR #84885.268), RT
ACN	Acetonitrile (VWR # 83642.320), RT
H_2O	Deionized water with a conductivity < 20 mSi (Millipore)
ESI Mass Calibrant	ESI-L low concentration running mix (Agilent G1969-85000), 4°C
Enzymes for digestions	Trypsin (Promega # V511A), Chymotrypsin (#V106A), Elastase (#V189A), Pepsin (#V195A), rLysC (#V167AQ), AspN (#V162A), GluC (#V165A); Stored according to manufacturer instructions
Column	Waters ACQUITY UPLC Peptide CSH C18, 130 Å, 1.7 µm, 150 mm x 2.1 mm
UHPLC	Agilent UHPLC 1290 Infinity
ESI-QTOF MS	impact II QTOF (Bruker)

Buffers	50 mM Ammonium bicarbonate pH 7.8 for Trypsin, AspN, GluC, rLysC digestions
	100 mM Tris-HCl, 10 mM $CaCl_2$, pH 8.0 for Chymotrypsin digestion
	40 mM HCl for Pepsin digestion
	50 mM Tris-HCl, pH 9.0 for Elastase digestion
	10 mM Sodium Phosphate Buffer (pH=7.4, RT) for PNGase F digestion

Sample preparation

- Six digest batches were prepared from a total of 30 µg of the antibody. The enzymes (Trypsin, Chymotrypsin, Elastase, Pepsin, Double digestion: AspN/LysC and GluC/LysC) were added to the sample aliquots with a ratio of 1:20 (w/w) separately, incubated for 4 h and 16 h at enzyme-specific temperatures. **Note**: Digest time may vary for different mAbs.
- Five µg (here, 5 µL) of the antibody were incubated with 0.5 µL PNGase F (1000 U/mL) in PBS buffer at 37°C for 1 h, for each proteolytic digestion.
- The deglycosylated sample (5 µL) was denatured with 14 µL 8 M urea and 1 µL 100 mM DTT at 50°C for 30 min, followed by spin down. Then 4 µL 100 mM IAA (freshly prepared) were added and the sample was incubated for 30 min in the dark, followed by spin down.
- Samples were diluted 5-fold with protease-specific buffers to a final concentration of urea <1 M.

- FA was added to a final concentration of 0.5% to terminate the digestion.

- For pepsin digestion, the deglycosylated sample was buffered with 40 mM HCl to a pH 1.0-3.0. Pepsin was added to the sample with an enzyme-to-substrate ratio of 1:20 (w/w). The mixture was incubated for 4 h or 16 h at 37°C. The digestion was terminated through heat denaturation (10 min, 95°C).

Data acquisition

All measurements were performed using a Bruker impact II QTOF mass spectrometer with ESI source, coupled to an Agilent 1290 Infinity UHPLC system. The LC-MS/MS-analyses were performed for each enzyme digest individually.

Data analysis

Spectra were automatically processed in PEAKS AB 2.0 software (Bioinformatic Solutions Inc.) for *de novo* sequencing: MS tolerance: 10 ppm; MS/MS tolerance: 0.05 Da; fixed modification: carbamidomethylation, variable modifications: oxidation, deamidation, pyro-Glu from Q and pyro-Glu from E. Subsequently, the peptide sequences were assembled into a single antibody sequence, which was manually validated by investigation of the MS/MS spectra to ensure correct sequence assignment.

Results and discussion

Antibody AF165 was digested with six enzymes or enzyme combinations separately according to the workflow displayed in Fig. 1. The combined datasets from six digests were submitted to PEAKS AB 2.0.

Assuming no prior knowledge of the protein sequence, AF165 data were *de novo* sequenced using the protein

HPLC peptide separation	
Column	Waters ACQUITY UPLC Peptide CSH C18, 130 Å, 1.7 µm, 150 mm x 2.1 mm
Mobile phase	Solvent A = 0.1% FA in H_2O Solvent B = 0.1% FA in ACN
Initial conditions	2% B at indicated temperature and flow
	The optimal flow rate and column temperature were determined experimentally to be 200 µL/min and 40°C
A 60-minute gradient (total 72-minute run time) was used	
Gradient:	5 min 2% B 65 min 40% B 66 min 90% B 68 min 90% B 69 min 2% B 72 min 2% B

MS data acquisition	
Method	AutoMS/MS top 20
	General: Positive mode, MS range 150-2200 m/z, Rolling Average Off, 0.5 Hz
	Source: 500 V, 4500 V, 2.0 Bar, 8.0 L/min N_2 flow, 220°C
	Transfer: Funnel1 400 Vpp, Funnel2 600 Vpp, isCID energy 0 eV, Hexapole RF 400 Vpp
	Quadrupole: ion energy 5 eV, set isolation mass 200 *m/z*
	Collision Cell: collision energy 10 eV, RF 1500 Vpp, transfer time 100 µs, prepulse storage time 10 µs
	0-5 min segment wo waste, 5-72 min segment to MS source
	Acquire line spectrum
Calibration	Instrument is calibrated using the Tuning Mix with a flow of 0.18 mL/h, Enhanced Quadratic

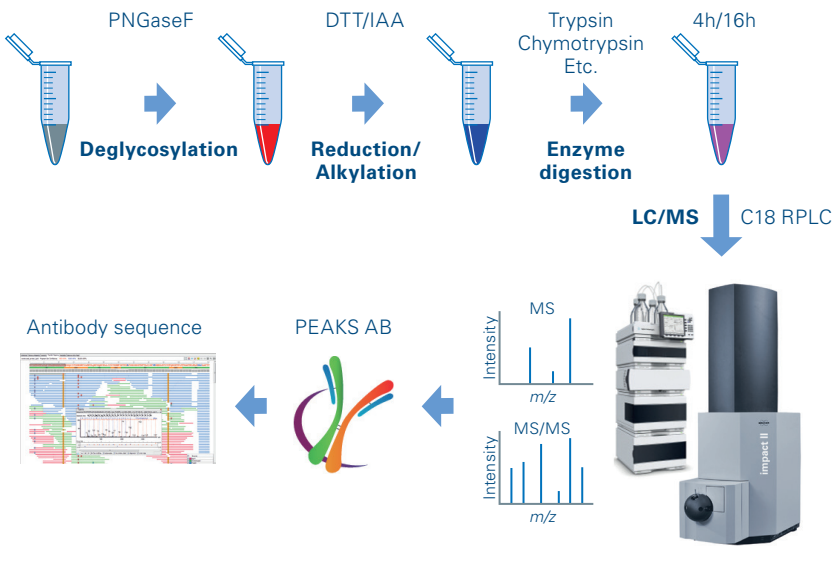


Figure 1: Schematic workflow description for antibody de novo *sequencing. 6 different enzyme digests were used.*

sequencing workflow in the PEAKS AB software [2]. Briefly, the spectra from all 6 digest datasets were sequenced directly to achieve peptide tags *de novo*. In parallel, the spectra were also submitted to sequence database searching; the *de novo* sequenced peptides were combined with the peptides obtained from database searching. All obtained peptide sequences were then processed with the De Bruijn graph method to form the final sequence with 100% sequence coverage of both light chain and heavy chain (Figs. 2, 3). An N-terminal pyro-Gln was also determined on the light chain. Spectra were manually inspected for confirmation of the assigned sequence. Based on the final sequences constructed and confidence levels reported by PEAKS AB (Fig. 4), further manual investigation allowed to verify and curate the sequencing result.

To resolve possible ambiguities of the isobaric leucine and isoleucine residues, PEAKS AB assigned them on the basis of homology statistics and the number of PSM (peptide-to-spectrum matches) obtained for the Pepsin and Chymotrypsin digestions (Figure 4).

In one example, the heavy chain sequence, HC [210-219] (KVDKKVEPKS), was insufficiently covered by MS/MS fragments to provide a conclusive sequence. However, the correct sequence was determined by the software for this region based on germline homology considerations and precursor ion information.

The true sequence of AF165 was then unblinded, and the experimentally *de novo* reconstructed sequence found to match 100%. All the CDR regions in both light chain and heavy chain were identified with 100% sequence match.

The wet lab workflow itself including LC-MS/MS data acquisition took about 4-5 days. Additional 3 days were required for data processing including manual evaluation and editing of the data with PEAKS AB software. In case of AF165, the light chain was 100% sequenced without further manual correction. However, a single GA sequence stretch of the heavy chain was misassigned as a Q, which was detected and corrected manually. In the heavy chain, HC[210-219] (KVDKKVEPKS) sequence ambiguities were resolved through homology search by PEAKS and confirmed manually using its accurate mass.

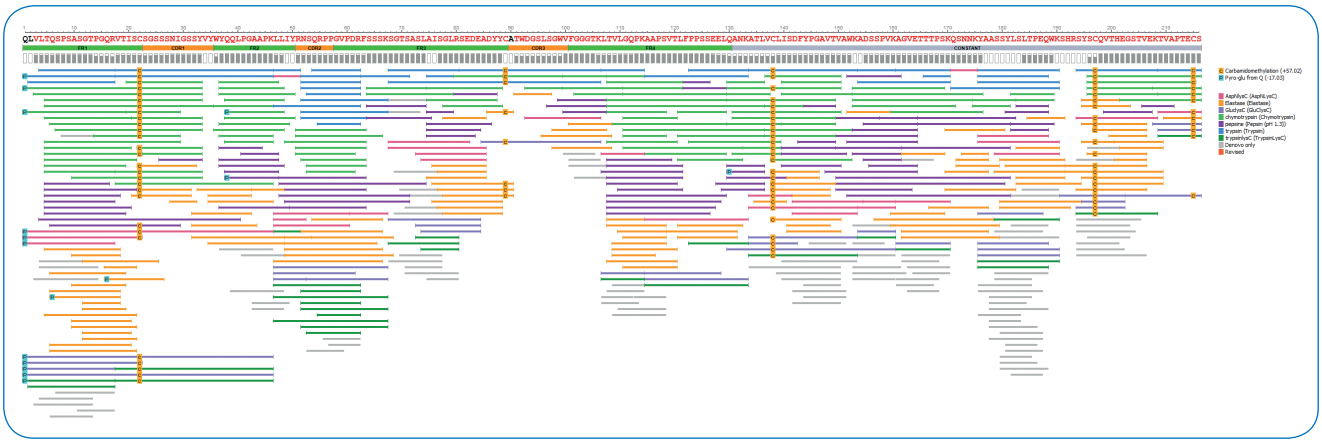


Figure 2: AF165 light chain amino acid sequence assembled by PEAKS AB after manual correction. Color codes indicate the peptide origin from various digests and critical processing steps.

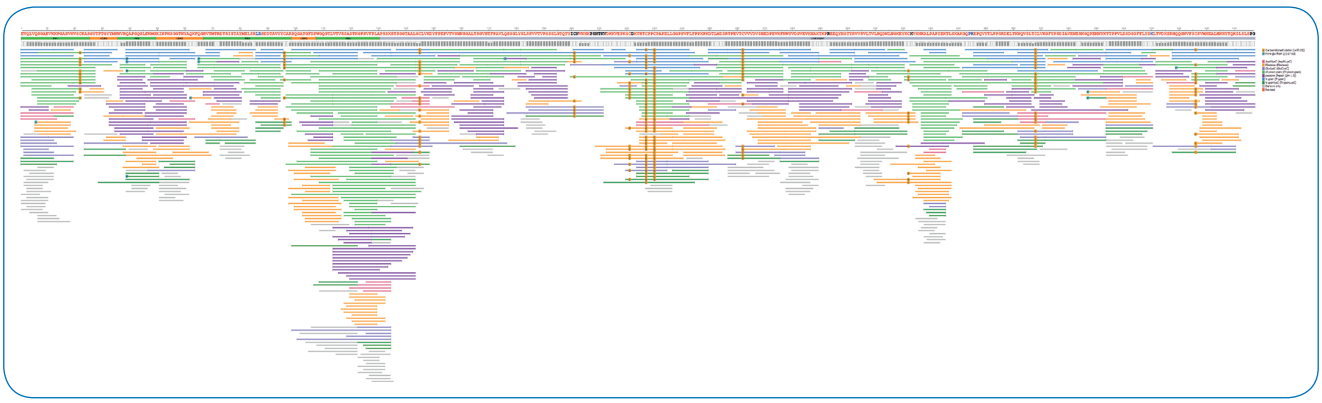
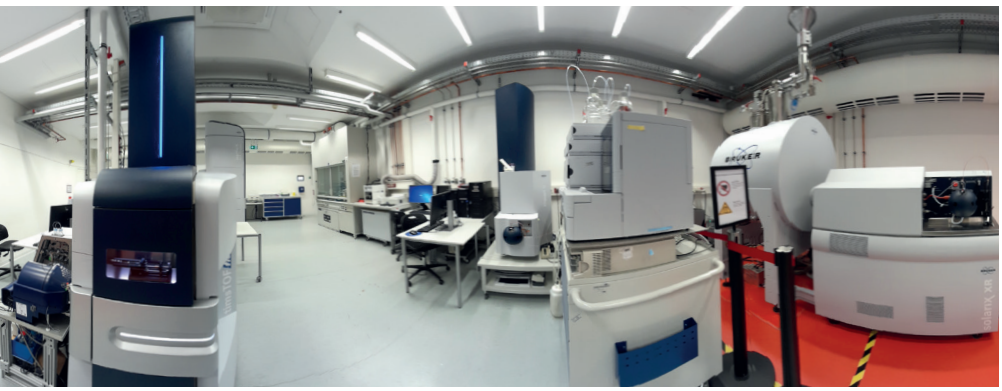


Figure 3: AF165 heavy chain amino acid sequence assembled by PEAKS AB after manual correction.

Leu/Ile Statistical Table

Region	Position	% of I	% of L	#PSMs induced by Chymotrypsin	#PSMs induced by Pepsin (pH 1.3)	Confidence
FR1	L@4	0	89	123	10	High
FR1	I@20	100	0	9	0	High
CDR1	I@29	94	1	0	0	High
FR2	L@40	1	73	0	1	Medium
FR2	L@47	0	95	0	195	High
FR2	L@48	0	95	226	43	High
FR2	I@49	94	1	0	19	Medium
FR3	L@74	0	100	312	81	High
FR3	I@76	99	0	0	0	High
FR3	L@79	0	100	12	2	High
CDR3	L@96	0	88	0	5	High
FR4	L@107	0	75	279	407	High
FR4	L@110	0	99	293	0	High
FR4	L@121	0	100	16	57	High
FR4	L@129	0	100	675	112	High
CONSTANT	L@136	0	100	176	41	High
CONSTANT	L@139	0	100	53	31	High
CONSTANT	I@140	99	0	0	0	High
CONSTANT	L@182	0	100	6	12	High
CONSTANT	L@184	0	100	0	0	Medium

Figure 4: Example of leucine and isoleucine differentiation on the light chain with PEAKS AB. In case of Ile-29, the homology statistics (94% Ile) and the absence of chymotrypsin and pepsin digestion at that site indicated the absence of Leu-29, hence the confidence of the amino acid at this position being Ile is high.



CeMOS mass spectrometry facility

Prof. Carsten Hopf (CeMOS): New de novo *sequencing capabilities on our CeMOS Q-TOF platform (www.cemos.org) utilizing Bruker impact II have opened up new avenues for research collaborations with and services for the biotech and biopharma industry.*

Conclusion

- A workflow is described for the full *de novo* sequencing analysis of an unknown monoclonal antibody based on multiple proteolytic digests, LC-MS/MS analysis on the impact II high-resolution QTOF and automatic peptide sequencing and sequence alignment using the PEAKS AB software. The workflow can be completed in approx. 4-8 days per monoclonal antibody, depending on the additional effort of manually curating lower confidence sequence patches. These are typically short sequence stretches for which low fragment ion coverages were obtained initially. Most ambiguities were automatically or semi-automatically resolved in Peaks AB using a database of known antibody sequences that allowed to correlate the results to the most likely known sequences.
- The workflow is suitable for routine antibody sequence determination. It provided the sequences of all 6 CDRs with great confidence and, in addition, determined pyro-Gln as LC terminal modification. The sequence obtained using the workflow was 100% correct.

CeMOS
Center for Mass Spectrometry and Optical Spectroscopy

INVIVO
InVivo BioTech Services GmbH
a BRUKER company

Learn More

You are looking for further Information?
Check out the link or scan the QR code.

www.bruker.com/massspectrometry

References

[1] Bandeira N, et al. (2008). *Automated* de novo *protein sequencing of monoclonal antibodies.* Nat Biotechnol. **26**(12):1336-8.

[2] Tran NH, et. al. (2016). *Complete* De Novo *Assembly of Monoclonal Antibody Sequences.* Scientific Reports, **6**:31730.

Bruker Daltonics GmbH & Co. KG

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com