Journal of proteome research

pubs.acs.org/jpr

Article

Complementary Methods for de Novo Monoclonal Antibody Sequencing to Achieve Complete Sequence Coverage

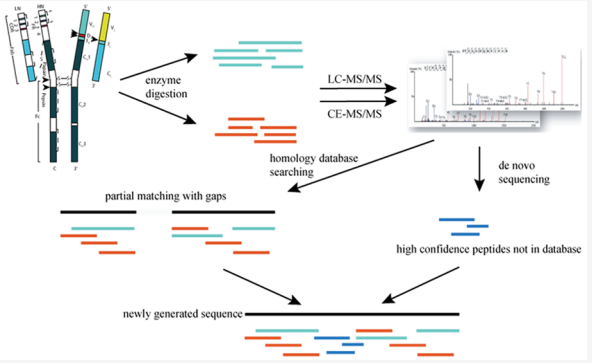
Jianhui Cheng,[§] Lingyu Wang,[§] Craig M. Rive,[§] Robert A. Holt, Gregg B. Morin,* and David D. Y. Chen*

Cite This: *J. Proteome Res.* 2020, 19, 2700−2707

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Mass spectrometry is a powerful tool for de novo sequencing of novel proteins. Recent efforts in this area have mainly focused on liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). Here, we present an alternative method, capillary electrophoresis tandem mass spectrometry (CE-MS/MS), for sequencing novel monoclonal antibodies. Using less than 200 ng in total of tryptic digest sample in a triplicated measurement, CE-MS/MS with pH-mediated focusing successfully sequenced mAb infliximab with 100% sequence coverage and 100% accuracy for the light chain and 96% coverage and 93% accuracy for the heavy chain. It was also demonstrated that CE-MS/MS gives comparable results, and in some cases, even better results, as compared to LC-MS/MS when used as a standalone technique. A combined workflow using both CE-MS/MS and LC-MS/MS was also used to sequence a novel antibody, anti-CD-176, resulting in the first proposed sequence for this mAb.



KEYWORDS: *novel monoclonal antibody, capillary electrophoresis, de novo sequencing and bottom-up proteomics, flow-through microvial interface, mass spectrometry*

Downloaded via UPPSALA UNIV on July 18, 2020 at 12:56:17 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

The development of monoclonal antibody (mAb)-based therapies was a major medical advance that improved the treatment of many conditions. Since 1985,[1] over 80 mAbs have been granted marketing approval as drugs, and the number of approvals is increasing rapidly.[2] However, systematic sequencing of mAbs is very challenging due to their variability.[3] In theory, each mAb is a unique protein that requires sequencing but has similar proteins in a homologous database.

Tandem mass spectrometry (MS/MS) is widely used in proteomics research to identify peptides/proteins by matching experimental mass spectra with databases.[4,5] However, this approach is less reliable when the target protein is novel. In particular, it often fails when applied to mAb sequencing because mAbs are not directly encoded by the genome,[6] and new sequences are constantly created. Until recently, the low-throughput Edman degradation protocol was the only viable method for mAb sequencing.[7] De novo protein sequencing, which is database independent, enables direct sequencing of novel proteins based on high-quality MS/MS spectra. Pioneering efforts in this area have resulted in the development of algorithms for de novo sequencing[8] that have been implemented in both free and commercial software packages (e.g., Quick Novor and PEAKS Studio).

LC-MS/MS has become a standard technology for sequencing novel proteins. The target protein is typically digested by multiple enzymes (separately or together), after which the resulting enzymatic peptides are separated by LC before being injected into a mass spectrometer. To improve the identification coverage, fragmentation methods such as collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), and electron-transfer dissociation (ETD) are commonly combined to generate as many fragments as possible. N. H. Tran et al. proposed an integrated assembly system which can automatically assemble full-length mAb sequences. Their work completely sequenced three mAbs of lengths 216−441 AA at 100% coverage and 96.64−100% accuracy. However, it focused on the development of assembly algorithms, not analytical methods. Other work has combined top-down or middle-down approaches with bottom-up approaches to improve identification coverage in mAb sequencing.[9,10] L. Fornelli et al. combined top-down and middle-down methods with the help of a state-of-the-art Orbitrap FT-MS equipped with multiple ion activation techniques and achieved overall sequence coverage (and post-translational-modification mapping) in line with that obtainable

Special Issue: Methods for Omics Research

Received: April 3, 2020
Published: April 27, 2020



ACS Publications

© 2020 American Chemical Society

by bottom-up MS strategies. Top-down and middle-down mass spectrometry offer reduced sample manipulation, limiting the risk of introducing artifactual modifications. However, their sequencing was based on a well-known mAb, rituximab.[11] This method is not yet able to sequence novel antibodies. Although most mAbs have mostly similar sequences with their homologies, it is very challenging to achieve 100% coverage for novel mAbs with high accuracy.

To our knowledge, LC-MS/MS has been the only tool used in novel mAb sequencing to date. It is well-known that CE provides complementary information to LC; it is being rapidly developed for bottom-up proteomics,[12,13] and is widely used in peptide mapping of mAbs,[14,15] intact proteins studies,[16−19] and antibody−drug conjugate analysis.[20] Here, we propose that, with careful optimization, CE-MS/MS, especially with the use of pH-mediated online focusing technics, could be a powerful alternative or complementary tool for sequencing novel mAbs.

METHODS

Chemicals and Materials

The anti-CD-176 monoclonal antibody was purchased from Thermo-Fisher Scientific (Nepean, ON, Canada). Infliximab in its medicinal formulation was provided by Agilent Technologies (Beijing, China). Acetonitrile, methanol, and formic acid were purchased from Thermo Fisher Scientific (Nepean, ON, Canada); all solvents were of analytical grade or better and were used without further purification. Ammonium acetate, Tris base, HCl, and NaOH were purchased from Sigma-Aldrich (St. Louis, MO); BCA Protein Assay Kits were purchased from Thermo Fisher Scientific (San Jose, CA). MS grade Trypsin, Glu-C, and Chymotrypsin were purchased from Thermo Fisher Scientific (Nepean, ON, Canada); hydroxyl propyl cellulose (HPC, molecular weight 100 000) was from Sigma-Aldrich (St. Louis, MO). Polyethylenimine (PEI, molecular weight 1500−1800) was purchased from Gelest Inc. (Morrisville, PA).

Sample Preparation

Original antibody (10 μg) in its medical formulation was exchanged into 100 mM Tris-HCl buffer solution (pH 7.8) through a centrifugal concentrator (molecular weight cutoff 10 000, BioSharp, Shanghai, China) three times to remove the medical excipients. The concentration of protein was measured using a BCA protein assay kit following the manufacturer's instructions. The sample was then resuspended in 8 M urea buffer and reduced in 5 mM dithiothreitol (DTT) at 37 °C for 40 min, after which 10 mM iodoacetamide (IAA) was added. This solution was incubated in the dark for 30 min at room temperature. Excess IAA was quenched by adding an additional 5 mM DTT and incubating for another 15 min at room temperature in the dark. The antibody sample was then diluted 1:4 with 100 mM Tris-HCl buffer to reduce the concentration of urea below 2 M, after which the antibody was digested by adding trypsin (for anti-CD 176 antibody, it was also digested by Glu-C or chymotrypsin) at an enzyme/protein ratio of 1:50 (w/w). The mixture was then incubated at 37 °C overnight, and the peptide mixture was desalted by a stage-tip method.[21] The desalted peptide mixture was stored in a freezer at −80 °C until use.

Capillary Coating

A positively charged capillary coated with PEI was prepared as follows: a 105 cm capillary (50 μm i.d. and 360 μm o.d., Polymicro Technologies, Phoenix, AZ) was sequentially rinsed with MeOH for 5 min, H_2O for 5 min, 0.1 M NaOH for 20 min, H_2O for 5 min, 0.1 M HCl for 20 min, H_2O for 5 min, and MeOH for 5 min and then dried by flushing with air for 10 min at a pressure of 20 psi. A 10% PEI solution in 80% methanol was then injected into the capillary at 30 psi for 9 min, after which the two outlets of the PEI-filled capillary were placed in a 1% F.A. solution for 720 min. The capillary was then rinsed with 1% F.A. for 5 min at 30 psi, and a −20 kV voltage was applied to test the stability of the current. Finally, the coated capillary was flushed with nitrogen and sealed for storage.

PEI coating was used in this experiment for the following reasons: (i) The PEI-coated capillary filled with the acidic separation background electrolyte (BGE) has positive charges on its inner surface, as do the proteolytic peptide analytes. The charge propulsion suppresses adsorption between the analytes and the capillary, thus eliminating possible peak tailing and broadening. (ii) As we adopted online pH barrage junction focusing to enhance the signal intensity in MS/MS data acquisition, a short alkaline pH barrage segment was injected into the separation capillary. As a covalently coated capillary, the PEI coating layer demonstrated sufficient chemical resistance to corrosion by the alkaline solution. The data analyzed in this study were acquired using a single PEI-coated capillary. (iii) In contrast to the neutrally coated capillary, the EOF in a PEI-coated capillary is reversed instead of being eliminated; therefore, the CE duration was reduced to <30 min while maintaining an acceptable signal window for mAbs analysis.

CE-MS/MS

Samples were analyzed using a PA800 CESI system (only the CE system was used, not its interface system, Beckman Coulter Inc., Fullerton, CA) connected to an Orbitrap Fusion Lumos Tribrid (Thermo Scientific) mass spectrometer via a homemade flow-through microvial interface.[22,23] The modifier solution used in this experiment consisted of 75% MeOH, 24.5% water, and 0.5% F.A. It was delivered by a pump at a flow rate of 400 nL/min through the interface to stabilize the electrospray.

To improve concentration sensitivity during CE, a fast and efficient pH-mediated sample stacking method called "pH barrage" that was previously developed in our group was used for online concentration.[24] The background electrolyte (BGE) for this method consisted of 200 mM F.A. in 50% MeOH (pH 2.2), and the peptides were dissolved in the BGE prior to analysis. Immediately before sample injection, a suitable quantity of a solution of 200 mM NH_3 and 20 mM F.A. in 50% MeOH (pH 10.2) was injected to create a pH barrage for the purpose of online stacking. When the peptides entered the high pH barrage region, ions contained fewer positive charges or even became negatively charged, causing them to move more rapidly toward the outlet. As a result, the peptides were concentrated between the boundaries of the pH barrage zone and the normal BGE zone. The optimized injection parameters for the barrage and sample plugs were 2 psi for 30 s and 2 psi for 60 s, respectively, and a 25.5 kV separation voltage was applied for 30 min. All CE parameters were controlled using the 32 Karat software package to ensure reproducibility.

The mass spectrometer was operated in data dependent acquisition (DDA) mode. Default parameters were used unless indicated below. Full MS^1 scans were acquired using the Orbitrap mass analyzer over a range of m/z 380−1500 with a resolution of 120 000 at 400 m/z. The AGC target was set to 4.0×10^5; the maximum injection time was 50 ms, and 1 microscan was performed. Top speed mode was used to select the most

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707

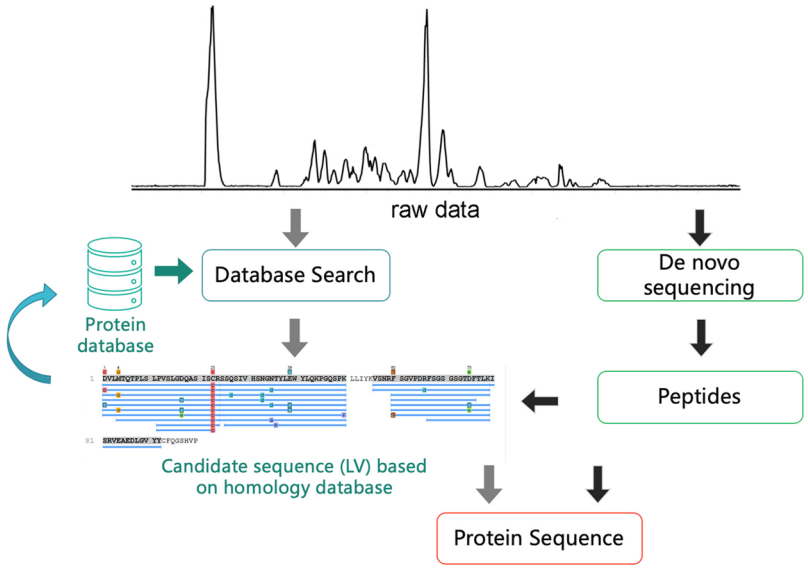


Figure 1. General workflow for sequencing of novel proteins via PEAKS based on an electropherogram. First, the obtained MS/MS spectra were searched against a public database of homologous proteins. The best matching homologous sequences were then modified based on the previous search results to generate a new database composed of the modified proteins. Meanwhile, the de novo peptides were inserted into the gaps to improve the coverage. This process was repeated until the sequence was either completed, or its coverage could not be improved any further based on the experimental data.

intense peaks with charge states of 2−7 for fragmentation. HCD was chosen as the activation type, with a normalized collision energy of 35%. MS^2 scans were acquired using the Orbitrap detector with a resolution of 30 000. For dynamic exclusion, the repeat count was set to 1, and the exclusion duration was 60 s; the mass tolerance was 10 ppm, and "exclude isotopes" was checked.

LC-MS/MS

Peptide mixtures were separated using an Ultimate 3000 RSLCnano system (Thermo Scientific) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with a nanospray ionization source. A self-packed LC column (Aqua 3 μm C18 125 Å, 100 μm (i.d.) × 15 cm (length)) was used in this experiment. The 42 min gradient (A, 0.1% F.A. in water; B, 0.1% F.A. in ACN) was as follows: 0−5 min, 2% B; 5−35 min, 2−30% B; 35−40 min, 30−45% B; 40−42 min, 45−90% B. A flow rate of 300 nL/min was used. The mass spectrometric settings were the same as those in the CE-based protocol.

Data Analysis

As shown in Figure 1, the MS/MS spectra obtained by LC-MS/MS or CE-MS/MS were searched against a homologous protein database. The cRAP contaminant database (ftp://ftp.thegpm.org/fasta/cRAP) containing 329 contaminant proteins was used to filter out peptides of interest from contaminant proteins. After the first round of searching, the best matched segments were carefully examined manually, especially those in the variable regions of both chains. High confidence segments (especially those in variable regions) were used for protein BLAST to generate another homology database. A second round of searching was then performed against the new BLAST-derived database. In most cases, this revealed a few candidate light and heavy chains with high coverage values. The well-matched peptide contigs were then examined carefully to identify possible ways of combining motifs from different candidates to generate a new target sequence with higher coverage. The de novo peptides were also carefully searched for candidate segments that fitted the candidate sequences. The database for third-round searching was generated based on the candidate sequences from the second-round searches. The second- and third-round search procedures were then repeated iteratively until the coverage of the sequence was completed, or its coverage could not be improved any further based on the experimental data.

Data analysis was performed with PEAKS Client (Version 7.0). During data refinement, scans were not filtered or merged, and precursors were corrected by mass only. The tolerance error for precursors was 20.0 ppm based on monoisotopic mass, while that for fragment ions was 0.02 Da. Enzymes were specified on a sample-by-sample basis. Carbamidomethylation (Cys) was set as a fixed modification, with oxidation (Met) and deamidation (Asn) as variable modifications. The results were filtered according to the following criteria:

(1) For peptides identified by database search, FDR was set at 1%.

(2) For peptides identified by de novo sequencing, the average local confidence (ALC) scores were required to be above 50. Further manual checks were performed for chosen candidates identified by the de novo algorithm.

RESULTS

Sequencing Infliximab with CE-MS/MS and LC-MS/MS

To evaluate the ability of CE-MS/MS to sequence unknown mAbs, we used it to sequence infliximab (sold under the brand name Remicade). Infliximab is a well-known chimeric mAb medication used to treat various autoimmune diseases. It has been studied extensively, and its sequence is included in most mAb databases. To simulate the workflow for sequencing novel mAbs, we removed all infliximab-related sequences from our database prior to data analysis.

The MS^2 spectra obtained using CE-MS/MS were searched against the homemade mAbs database. A high confidence peptide from the light chain variable region (LLIKYASESMSGIPSR) was used for protein BLAST to generate a BLAST-

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707

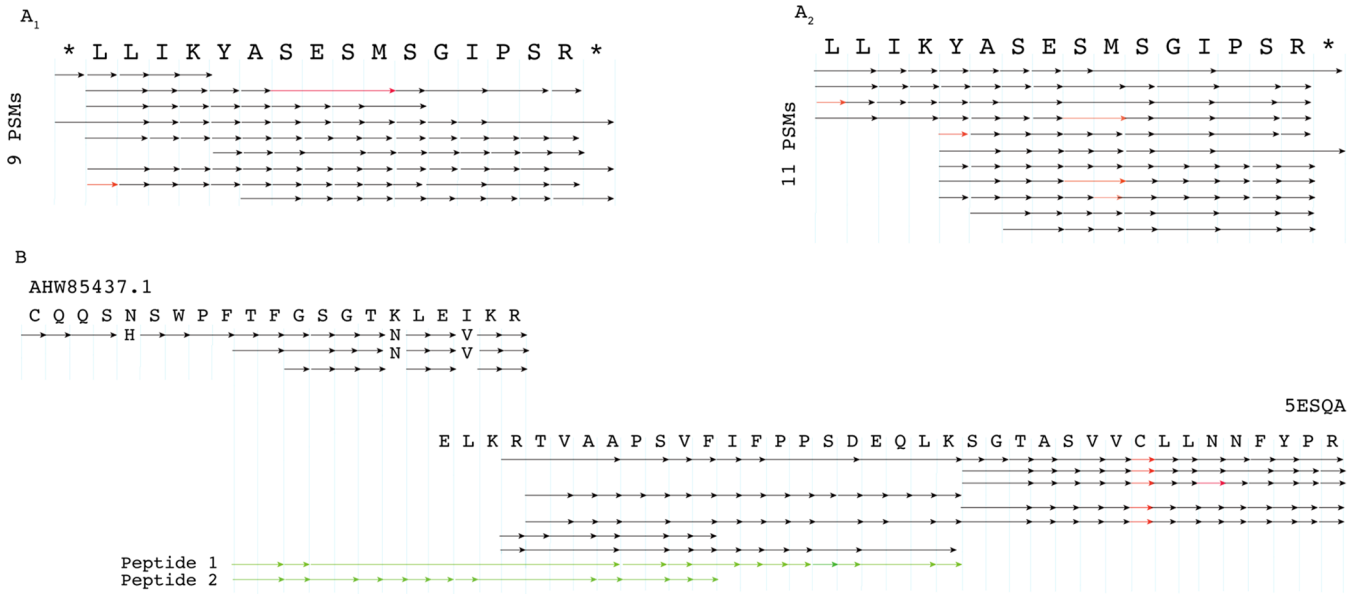


Figure 2. (A) Peptide LLIKYASESMSGIPSR identified in the first-round homologous database search. Parts A_1 and A_2 show peptide contigs supported by 9 PSMs obtained using CE-MS/MS and 11 PSMs obtained using LC-MS/MS, respectively. (B) Assembly of a long peptide confirmed by the overlapping amino acid sequences of two new peptide contigs. The subsequent identification of peptides 1 and 2 against the assembled protein confirmed the accuracy of the assembly. Each PSM used in the assembly is represented by a horizontal line with a set of superimposed arrows indicating its sequence of covered masses; deviations from the consensus sequence are indicated by red arrows.

1–60	DILLTQSPAI	LSVSPGERVS	FSCRASQFVG	SSIHWYQQRT	NGSPRLLIKY ASESMSGIPS
61–120	RFSGSGSGTD	FTLSINTVES	EDIADYYCQQ	SHSWPFTFGS	GTNLEVKRTV AAPSVFIFPP
121–180	SDEQLKSGTA	SVVCLLNNFY	PREAKVQWKV	DNALQSGNSQ	ESVTEQDSKD STYSLSSTLT
181–214	LSKADYEKHK	VYACEVTHQG	LSSPVTKSFN	RGEC	

1–60	EVKLEESGGG	LVQPGGSMKL	SCVASGFLFS	NHWMNWVRQS	PEKGLEWVAE IRSKSLNSAT
61–120	HYAESVKGRF	TISRDDSKGT	VYLQMTDLRT	EDTGVYYCSR	NYYGAMHSWW GEGTTLTVSS
121–180	ASTKGPSVFP	LAPSSKSTSG	GTAALGCLVK	**DYFPEPVTVS**	**W**NSGALTSGV HTFPAVLQSS
181–240	GLYSLSSVVT	VPSSSLGTKT	YLCNVNHKPS	NTKVDKRVES	KYGPPCPPCP APPVAGGPSV
241–300	FLFPPKPKDT	LMISRTPEVT	CVVVDVSHED	PEVQFNWYVD	GVEVHNAK**TK** **PR**EEQFNSTY
301–360	RVVSVLTVLH	QDWLNGKEYK	CKVSNKALPA	PIEKTISKAK	GQPREPQVYT LPPSQEEMTK
361–420	NQVSLTCLVK	GFYPSDIAVE	WESNGQPENN	YKTTPPVLDS	DGSFFLYSKL TVDKSRWQQG
421–447	NVFSCSVMHE	ALHNHYTQK**S**	**LSLSPGK**		

Figure 3. Assembled light and heavy chain sequences of infliximab based on data acquired by CE-MS/MS. Residues in this sequence that differ from those in the true sequence are shown in red (for further discussion, see Figure S14 in the Supporting Information), and uncovered sequences are in bold.

derived database. Figure 2A shows the peptide-spectrum matchings (PSMs) obtained by LC-MS/MS and CE-MS/MS that support the presence of this peptide contig in the target mAb. After the second round of searching against the BLAST-derived database, two proteins 5ESQ_A and AHW85743.1 (seq0_L and seq1_L, respectively; in seq0_L CQQSNSWPFTFGSGTKLEIKR was corrected to be CQQSHSWPFTFGSGTNLEVKR, see the Supporting Information for details) with high coverages at different regions were combined to generate a new candidate sequence with a few gaps for the third round of searching. This process is illustrated in Figure 2B. Before the start of the third round of searching, some modifications were made to the candidate light chain sequence based on the peptides identified by de novo sequencing. The newly generated light chain candidates were used as a database for the third round of searching. The best-matched candidate, whose sequence is shown in Figure 3, was taken as the final light chain. Its sequence was identical to the true infliximab light chain. In other words, using only 3 × 50 ng samples of a tryptic

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707

Light chain

1-70 DVLMTQTPLS LPVSLGDQAS ISCRSSQSIV HSNGNTYLEW YLQKPGQSPK LLIYKVSNRF SGVPDRFSGS

71-140 GSGTDFTLKI SRVEAEDLGV YYCFQGSHVP YTFGGGTKLE IKRADAAPTV SIFPPSSEQL TSGGASVVCF

141-210 LNNFYPKDIN VKWKIDGSER QNGVLNSWTD QDSKDSTYSM SSTLTLTKDE YERHNSYTCE ATHKTSTSPI

211-219 VKSFNRNEC

Heavy chain

1-70 QVQLQQSGAE LVRPGTSVKI SCKASGYTFT NYWIGWVKQR PGHGLEWIGD IYPGGGYTNY NEKFKGKATL

71-140 TADTSSSTAY MQLSSLTSED SAVYYCFQGN KFLRWGAGTF AHNVTVTRES QSFPNVFPLV SCESPLSDKN

141-210 LVAMGCLARD FLPSTISFTW NYQNNTEVIQ GIRTFPTLRT GGKYLATSQV LLSPKSILEG SDEYLVCKIH

211-280 YGGKNRDLHV PIPAVAEMNP NVNVFVPPRD GFSGPAPRKS KLICEATNFT PKPITVSWLK DGKLVESGFT

281-350 TDPVTIENKG STPQTYKVIS TLTISEIDWL NLNVYTCRVD HRGLTFLKNV SSTCAASPST DILTFTIPPS

351-420 FADIFLSK**SA NLTCLVSNLA TYE**TLNISWA SQSGEPLETK IKIMESHPNG TFSKAGVASV CVEDWNNRKE

421-490 FVCTVTHRDL PSPQKKFISK PNEVHKHPPA VYLLPPAREQ LNLRESATVT CLVKGFSPAD ISVQWLQRGQ

491-560 LLPQEKYVTS APMPEPGAPG FYFTHSILTV TEEEWNSGET YTCVVGHEAL PHLVTERTVD KSTGKPTLYN

561-573 VSLIMSDTGG TCY

Figure 4. Assembled amino acid sequence of anti-CD-176: a light chain with an average mass of 24 158 composed of 219 amino acids and a heavy chain with an average mass of 63 119 composed 573 amino acids. The assembled low confidence sequence containing a possible CDR3 (red letters) in this experiment is circled in red. The bold amino acids were not covered.

infliximab digest, CE-MS/MS was able to fully sequence its light chain with 100% accuracy.

The heavy chain of infliximab has 447 amino acids (AAs), more than twice as many as the light chain, which made the assembly more difficult. The initial homology database search revealed several segments with high confidence. Two of these segments—the peptide contigs QSPEKGLEWVAEIR located in the variable region and VVSVLTVHQDWLNGK located in the constant region—were used for protein BLAST to generate the BLAST-derived database. After several rounds of searching and database modification, the candidate heavy chain sequence shown in Figure 3 was obtained. Residues in this sequence that differ from those in the true sequence are shown in red. In this case, the CE-MS/MS-based sequencing strategy achieved 95% coverage with an accuracy of 402/424 AAs. It should be pointed out that the protocol used in this work cannot yet distinguish between leucine and isoleucine, and I-to-L substitutions were not considered as mismatching.

The infliximab digest was also analyzed by LC-MS/MS. The sample loading in this case was around 1 μg per LC run. The search results based on LC-MS/MS data filled the missing tail "SLSLSPGK", which brought us closer to the true sequence.

Sequencing Anti-CD-176 with CE-MS/MS and LC-MS/MS

CD-176 (the Thomsen-Friedenreich antigen) is a tumor-associated carbohydrate structure that is expressed weakly in normal adult tissues but exhibits fatal levels of expression in hyperplasia or malignancy. The mAb anti-CD-176 can induce apoptosis of CD-176-positive leukemic cells and is widely used in CD-176-associated signaling pathway analysis,[25,26] but its sequence is currently unknown.

Three proteases (trypsin, Glu-C, and chymotrypsin) were used to generate different peptide mixtures from anti-CD-176 in order to produce large numbers of peptides with overlapping amino acid sequences to facilitate sequence assembly. Each sample was then analyzed by LC-MS/MS and CE-MS/MS. The sample loading was about 1 μg for LC analysis and 100 ng for CE analysis.

As when sequencing infliximab, we initiated the sequencing of anti-CD-176 by performing homologous database searches. The first round of searching yielded a high-confidence light chain candidate with a gap after Y-92. Two peptides from the de novo sequencing process, YYCFQHNKFLR and YYCFQGSHVPY, were inserted into the candidate sequence to generate two light chain candidates for the next round of searching. Because the coverage achieved during the first round was 97% (only the complementary determining region 3 (CDR3) in the variable region was still missing), the protein BLAST step for the light chain was skipped. After the second round of database searching against the two new candidates, YYCFQGSHVPY was identified as the missing gap motif because of its multiple overlaps with the surrounding sequences. A light chain sequence comprising 219 amino acids was ultimately assembled, as shown in Figure 4. Interestingly, all of the peptide contigs supporting the light chain tail sequence ATHKTSTSPIVKSFNRNEC were only detected by CE-MS/MS.

The high confidence peptide QRPGHGLEWIGDIYPGGGY-TNYNEKFK from the variable region and GVASVC-VEDWNNRKEFVCTVTHRDLPSPQKK (IgM C3) from the constant region were used for protein BLAST to generate the BLAST database for the anti-CD-176 heavy chain. During the second round of searching, the previously determined light

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707

chain sequence was added to the contaminant database. After repeated database searching and careful assembly, a finalized anti-CD-176 heavy chain with a coverage of 97% was generated, as shown in Figure 4. The search process also revealed the possible presence of other antibodies in the sample. To identify these antibodies, the high confidence peptides APQVYVLPPPEEEMTKK and RSVSELPIMHQDWLNGKEFK were searched by BLAST against a nonredundant protein database (nr) to generate a BLAST database. Careful analysis of the resulting data suggested that (excluding IgM) the sample contained heavy constant regions from at least two IgG antibodies. No other high confidence heavy chain variable regions were assembled based on our data (see the Supporting Information for details).

Sequencing the highly variable CDR3 regions of mAbs is very challenging, and the coexistence of bovine serum albumin (BSA, at a concentration 10 times that of mAb) with anti-CD-176 and the other antibodies in the sample created highly complex MS^2 spectra, making interpretation even more difficult. Three de novo peptide sequences YYCFQHNKFLR, RWGAGTF, and FAHNVTVTR, which had overlapping amino acids or similarity to sequences in the database, were used to generate a candidate sequence containing a possible CDR3: QGNKFLR, in the order shown with one residue overlap (Figure 4). It should be noted that this CDR3 assembly was of low confidence due to a lack of multiple overlapping subsequences. Confidence in the sequence could be increased by repeating the analysis with prepurified samples or using CE-MS/MS in conjunction with a top-/middle-down proteomics method.

DISCUSSION

To validate the complementarity of CE to LC, we first evaluated the ability of our CE-MS/MS method to identify infliximab. Unlike in the sequencing process above, we used the known sequence of infliximab in the database search.

Table 1 compares the identified peptides and protein coverages for our CE-based method to those achieved with LC-MS/MS. In general, CE outperformed LC with 10 times less sample consumption. In more detail, with 1 μg sample injection, LC identified 357 peptides from the heavy chain with a coverage of 85% and 475 peptides from light chain with a coverage of 100%. In comparison, CE consumed about 0.15 μg of sample for each run and identified 413 peptides from the heavy chain, and 425 peptides from the light chain. In addition, the coverage for the heavy chain was increased to 95%. When results from CE and LC were combined, the coverage of the heavy chain reached 97%.

Table 1. Sample Consumption, Numbers of Identified Peptides, And Sequence Coverages for LC-MS/MS and CE-MS/MS of Infliximab

		heavy chain		light chain	
method	total sample consumption (ng)	no. peptides	coverage	no. peptides	coverage
LC-MS/MS	~3000	357	85%	475	100%
CE-MS/MS	~150	413	95%	425	100%

It should be noted that our group has worked with CE-MS for a long time, and there might be concerns that the better coverage by CE-MS is simply the results of our preference of the technology. An analysis of the MW distribution of the identified peptides revealed that CE-MS/MS outperformed LC-MS/MS with respect to the identification of low MW peptides and also with respect to peptides of high MW (>6k), as shown in Figure 5B. This is due to the separation mechanism of reversed phase LC (the LC method used in this study): it is suboptimal for both small and very large peptides, which are either not retained on the columns or retained irreversibly. This is one of the reasons we sought to develop a CE-based mAb sequencing method.

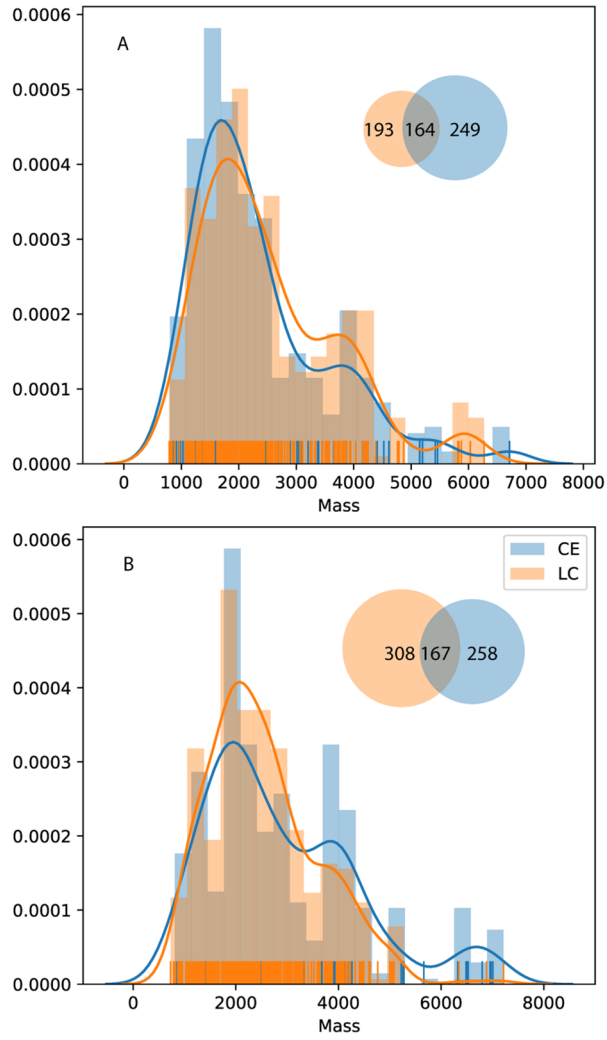


Figure 5. Kernel density estimations of the peptides' molecular weight distributions (A, heavy chain; B, light chain) and Venn diagrams showing the overlap of peptides identified by CE-MS/MS and LC-MS/MS in the infliximab digest (inserted subfigures).

To further investigate the complementarity of the CE-based method to LC, we also analyzed the results obtained from anti-CD 176 sequencing. The final sequence obtained in this study was used as the search database for this discussion. 89 peptides from the heavy chain were identified by CE-MS/MS from tryptic digests of anti-CD 176 mAb, while LC-MS/MS identified 56 peptides, 27 of which were also identified by CE-MS/MS. The numbers of peptides detected by CE after Glu-C and chymotryptic digests were also compared with LC results, as shown in the Venn diagrams inserted in Figure 6.

The MW distributions of the identified peptides were also studied. The results were in agreement with our previous conclusion, that the CE-based method outperformed LC at

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707

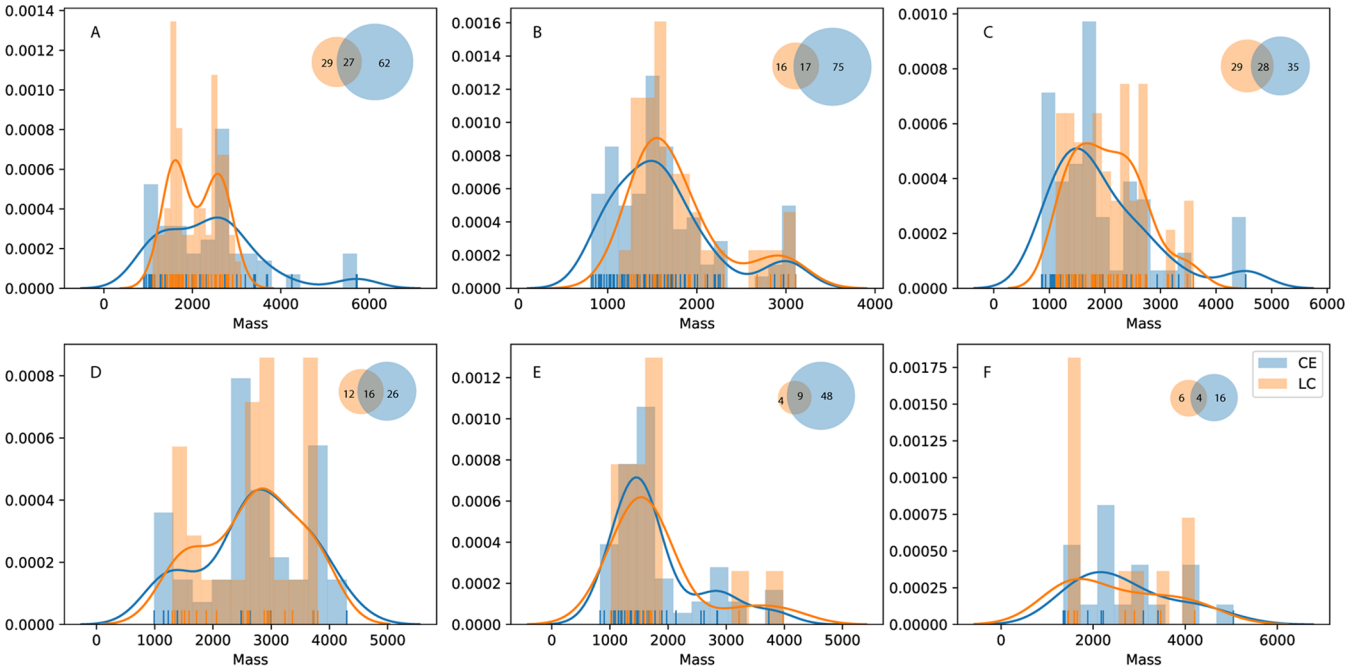


Figure 6. Kernel density estimations of anti-CD 176 peptides' molecular weight distributions: Parts A and D show results for tryptic digests of heavy and light chains; parts B and E show results for chymotryptic digests of heavy and light chains; parts C and F show results for Glu-C digests of heavy and light chains. The corresponding Venn diagrams showing the overlap of peptides identified by CE-MS/MS and LC-MS/MS are inserted as subfigures.

detecting low MW peptides (<1.5 k) in most cases, providing highly complementary information for peptide mapping. This conclusion was further confirmed by a study based on BSA mapping; see the SI for further information.

CONCLUSION

A CE-MS/MS workflow was developed and was shown to be complementary to LC-MS/MS workflows for sequencing novel mAbs. A homemade microvial flow-through interface was used to connect a CE system to a Thermo Scientific Orbitrap Fusion Lumos mass spectrometer for peptide identification. The use of CE-MS/MS as an alternative to LC-MS/MS has at least two clear benefits. First, each CE-MS/MS run took less than 30 min, making it much less time-consuming than traditional LC-MS/MS, which typically takes 50 min per run including the time needed for analytical column preconditioning. Second, individual CE runs required only 50−100 ng of sample, meaning that residual sample material left behind after LC injection can be used to generate a large number of complementary peptide identifications. This latter point also means that a novel antibody could potentially be sequenced using only about 200 ng of sampled material if a purely CE-based workflow were adopted. These results demonstrate that CE-MS/MS is perfectly suited to mAb sequencing because of its efficiency, sensitivity, and complementarity to LC-MS/MS.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00223.

Matched contigs for infliximab, heavy and light chain candidates, part of the final heavy chain variable region candidate for infliximab; all the peptide contigs for the C-terminal region of the anti-CD-176 light chain, two contaminating IgG heavy constant regions, finalized light chain CDR3 for anti-CD-176 (FQGSHVPYT) and its neighboring regions, low confidence assembled heavy chain CDR3 for heavy chain anti-CD-176, Venn diagrams, and a comparison between the true sequence of the infliximab heavy chain (the template) and the sequence obtained in this study (PDF)

AUTHOR INFORMATION

Corresponding Authors

Gregg B. Morin − *Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, University of British Columbia, Vancouver, British Columbia V5Z 1L3, Canada;* orcid.org/0000-0001-8949-4374; Email: gmorin@bcgsc.ca

David D. Y. Chen − *Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada;* orcid.org/0000-0002-3669-6041; Email: chen@chem.ubc.ca

Authors

Jianhui Cheng − *Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*

Lingyu Wang − *Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*

Craig M. Rive − *Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, University of British Columbia, Vancouver, British Columbia V5Z 1L3, Canada*

Robert A. Holt − *Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, University of British Columbia, Vancouver, British Columbia V5Z 1L3, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00223

Author Contributions

§J.C., L.W., and C.M.R. contributed equally to this work.

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, and Mitacs Canada Accelerate Cluster Program sponsored by Lipont Pharmaceuticals Inc., Richmond, BC, Canada. The mass spectrometric analysis was performed in Nanjing Normal University, China.

■ REFERENCES

(1) Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228* (4705), 1315−1317.
(2) Kaplon, H.; Reichert, J. M. Antibodies to Watch in 2019. *mAbs* **2019**, *11* (2), 219−238.
(3) 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*, 31730.
(4) Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314* (5796), 65−66.
(5) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198−207.
(6) Tonegawa, S. Somatic Generation of Antibody Diversity. *Nature* **1983**, *302* (5909), 575−581.
(7) Pham, V.; Henzel, W. J.; Arnott, D.; Hymowitz, S.; Sandoval, W. N.; Truong, B. T.; Lowman, H.; Lill, J. R. De Novo Proteomic Sequencing of a Monoclonal Antibody Raised against OX40 Ligand. *Anal. Biochem.* **2006**, *352* (1), 77−86.
(8) Muth, T.; Renard, B. Y. Evaluating de Novo Sequencing in Proteomics: Already an Accurate Alternative to Database-Driven Peptide Identification? *Briefings Bioinf.* **2018**, *19* (5), 954−970.
(9) Cotham, V. C.; Horton, A. P.; Lee, J.; Georgiou, G.; Brodbelt, J. S. Middle-Down 193-Nm Ultraviolet Photodissociation for Unambiguous Antibody Identification and Its Implications for Immunoproteomic Analysis. *Anal. Chem.* **2017**, *89* (12), 6498−6504.
(10) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2014**, *86* (6), 3005−3012.
(11) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90* (14), 8421−8429.
(12) Sun, L.; Zhu, G.; Zhao, Y.; Yan, X.; Mou, S.; Dovichi, N. J. Ultrasensitive and Fast Bottom-up Analysis of Femtogram Amounts of Complex Proteome Digests. *Angew. Chem., Int. Ed.* **2013**, *52* (51), 13661−13664.
(13) Sun, L.; Hebert, A. S.; Yan, X.; Zhao, Y.; Westphall, M. S.; Rush, M. J. P.; Zhu, G.; Champion, M. M.; Coon, J. J.; Dovichi, N. J. Over 10 000 Peptide Identifications from the HeLa Proteome by Using Single-Shot Capillary Zone Electrophoresis Combined with Tandem Mass Spectrometry. *Angew. Chem.* **2014**, *126* (50), 14151−14153.
(14) Gahoual, R.; Burr, A.; Busnel, J.-M.; Kuhn, L.; Hammann, P.; Beck, A.; François, Y.-N.; Leize-Wagner, E. Rapid and Multi-Level Characterization of Trastuzumab Using Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry. *mAbs* **2013**, *5* (3), 479−490.
(15) Dada, O. O.; Zhao, Y.; Jaya, N.; Salas-Solano, O. High-Resolution Capillary Zone Electrophoresis with Mass Spectrometry Peptide Mapping of Therapeutic Proteins: Improved Separation with Mixed Aqueous−Aprotic Dipolar Solvents (N,N-Dimethylacetamide and N,N-Dimethylformamide) as the Background Electrolyte. *Anal. Chem.* **2017**, *89* (21), 11227−11235.
(16) Redman, E. A.; Batz, N. G.; Mellors, J. S.; Ramsey, J. M. Integrated Microfluidic Capillary Electrophoresis-Electrospray Ionization Devices with Online MS Detection for the Separation and Characterization of Intact Monoclonal Antibody Variants. *Anal. Chem.* **2015**, *87* (4), 2264−2272.
(17) Shen, Y.; Zhao, X.; Wang, G.; Chen, D. D. Y. Differential Hydrogen/Deuterium Exchange during Proteoform Separation Enables Characterization of Conformational Differences between Coexisting Protein States. *Anal. Chem.* **2019**, *91* (6), 3805−3809.
(18) Wang, L.; Bo, T.; Zhang, Z.; Wang, G.; Tong, W.; Da Yong Chen, D. High Resolution Capillary Isoelectric Focusing Mass Spectrometry Analysis of Peptides, Proteins, And Monoclonal Antibodies with a Flow-through Microvial Interface. *Anal. Chem.* **2018**, *90* (15), 9495−9503.
(19) Wang, L.; Chen, D. D. Y. Analysis of Four Therapeutic Monoclonal Antibodies by Online Capillary Isoelectric Focusing Directly Coupled to Quadrupole Time-of-Flight Mass Spectrometry. *Electrophoresis* **2019**, *40* (21), 2899−2907.
(20) Dada, O. O.; Zhao, Y.; Jaya, N.; Salas-Solano, O. High-Resolution Capillary Zone Electrophoresis with Mass Spectrometry Peptide Mapping of Therapeutic Proteins: Peptide Recovery and Post-Translational Modification Analysis in Monoclonal Antibodies and Antibody−Drug Conjugates. *Anal. Chem.* **2017**, *89* (21), 11236−11242.
(21) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **2003**, *75* (3), 663−670.
(22) Zhong, X.; Maxwell, E. J.; Chen, D. D. Y. Mass Transport in a Micro Flow-Through Vial of a Junction-at-the-Tip Capillary Electrophoresis-Mass Spectrometry Interface. *Anal. Chem.* **2011**, *83* (12), 4916−4923.
(23) Maxwell, E. J.; Zhong, X.; Chen, D. D. Y. Asymmetrical Emitter Geometries for Increased Range of Stable Electrospray Flow Rates. *Anal. Chem.* **2010**, *82* (20), 8377−8381.
(24) Wang, L.; Tong, W.; Chen, D. D. Y. Optimization of Dynamic PH Barrage Junction Focusing for Weakly Alkaline or Zwitterionic Analytes in Capillary Electrophoresis. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2018**, *1095*, 94−102.
(25) Yi, B.; Zhang, M.; Schwartz-Albiez, R.; Cao, Y. Mechanisms of the Apoptosis Induced by CD176 Antibody in Human Leukemic Cells. *Int. J. Oncol.* **2011**, *38* (6), 1565−1573.
(26) Pérez, J. de J.; Udeshi, N. D.; Shabanowitz, J.; Ciordia, S.; Juárez, S.; Scott, C. L.; Olszewski, N. E.; Hunt, D. F.; García, J. A. O-GlcNAc Modification of the Coat Protein of the Potyvirus Plum Pox Virus Enhances Viral Infection. *Virology* **2013**, *442* (2), 122−131.

https://dx.doi.org/10.1021/acs.jproteome.0c00223
J. Proteome Res. 2020, 19, 2700−2707