Protein Sequencing Research Group (PSRG): Results of the PSRG 2010 Study: Edman and Mass Spectrometric Terminal Sequencing of a Monoclonal Antibody

J.S.Smith [1], W.Sandoval [2] ,P.Hunziker [3] ,B.Xiang [4] ,J.Walters [5], K.G.Mawuenyega [6], J.Simpson [7]

[1]University of Texas Medical Branch, Galveston, TX, United States, [2]Genentech, Inc., South San Francisco, CA, United States, [3]University of Zurich, Zurich, Switzerland, [4]Monsanto Company, St. Louis, MO, United States, [5]Sigma-Aldrich, St. Louis, MO, United States, [6]Washington University School of Medicine, St. Louis, MO, United States, [7]SAIC/National Cancer Institute, Frederick, MD, United States

Introduction:

N-terminal sequence analysis is an indispensable bioanalytical tool in the protein chemistry laboratory. N-terminal analysis is necessary for the quality control of protein biologics, for determining sites of biologically relevant proteolytic cleavage events, and is vital for the *de novo* characterization of monoclonal antibodies.[1] Automated Edman degradation has historically been the method of choice for these analyses, though in recent years alternative mass spectrometric methods for determining N-terminal sequence have surfaced[2,3]. The PSRG 2009 study made comparisons between the established Edman degradation techniques and alternative MS based techniques. Both methodologies were able to obtain N-terminal sequence information for the proteins provided however the study indicated reliance on protein databases to obtain N-terminal data from the MS technology.

For the 2010 study, the PSRG distributed an intact monoclonal antibody. The variable regions for the heavy and light chains, which include the N-termini, are not recorded in the protein databases. The complexity of this sample challenged participants with a molecule containing two N-termini; one of which was N-terminally blocked. The analysis of the results of the 2010 study focuses on the ability of the participants to obtain terminal sequence, with a comparison of the results highlighting the differences evident between classical Edman chemistry and alternative (MS based) technologies. Information on the type of instruments and protocols are also reported.

Study Methods:

The PSRG distributed a commercially available monoclonal antibody to 46 participants requesting the sample. The original commercial preparation of the antibody contained glycerol. Prior to distribution, the glycerol was removed with a chloroform/methanol precipitation and 50 µg of the resulting antibody was aliquoted in 1.5 ml tubes. Briefly, 50 µl of antibody, at 1 µg/µl, was placed in 1.5 ml tubes. 150 µl of methanol was added to each tube and vortexed. This was followed by the addition of 50 µl chloroform and 200 µl water with vortexing between the additions. The mixture was centrifuged at 14000 x g for 5 minutes. The top (aqueous) layer was carefully removed and 200 µl of methanol was added. The tube was vortexed and centrifuged again for 5 minutes at 14000 x g. The supernatant was removed and the resulting antibody pellet was dried in a speed vac. Two identical tubes containing 50 µg of antibody were mailed to each participant. Each participant requesting a sample was issued an anonymous code number. The participants were asked to use their code number to report their study data using a Google Docs web site.

Typical Participants Methods

Edman Degradation Method

Light chain:

•Most participants separated the antibody chains by SDS-PAGE and electroblotted to PVDF for sequencing. Others performed HPLC and loaded samples onto a TFA treated GFF support. In some cases sample was loaded directly onto PVDF or GFF and sequenced.

Heavy chain:

•Chains were separated on SDS-PAGE and blotted to PVDF for Edman degradation.

•For labs who pursued sequence determination after determining the heavy chain was blocked, a de-blocking step was performed either in solution or on membrane using *pyroglutamate aminopeptidase* (PGAP)

Bottom-up MS Method

- Protein separation using electrophoresis or HPLC.
- Protein digestion by enzyme.
- ESI-MS or MALDI-MS.
- ESI-MSMS or MALDI-MSMS.
- Protein identification through database searching.
- Protein terminal peptide identification or sequencing.

Top Down MS Method

- Heavy and light chains were separated by HPLC after reduction (and optional alkylation) of the intact antibody.
- Masses of the intact heavy and light chains determined by MALDI-TOF using sinnapinic or dihydroxybenzoic acid.
- Sequencing of heavy and light chain by in-source decay combined with MS/MS data obtained from fragments produced by in-source decay using 2,5-diaminonaphtalene (DAN) or 2,5-dihydrobenzioc acid (DHB) as matrix.

Instrumentation and Methods:

Edman

Lab #	Instrument	Reconstitution Buffer	Separation	Deblocking	Methods
PSRG-1	ABI Procise 494 HT	250mM phosphate, pH7.0, 50mM DTT, 5mMEDTA	SDS-PAGE	PGAP	De-blocked, SDS-PAGE, blot to PVDF
PSRG-2	ABI Procise 494 HT	0.1M AmBic	SDS-PAGE	PGAP	Reduced (DTT), alkylated (NIPIA), de-blocked, SDS-PAGE, blot to PVDF
PSRG-3	N/A	25 mm AmBic	HPLC, gel filtration	none	Did not recover sample after HPLC
3	ABI Procise 494	100mM TRIS then SDS sample buffer	SDS-PAGE	PGAP	SDS-PAGE, blot to PVDF, bands sequenced
34	ABI Procise 494 HT	8M Guanidine	HPLC	None	Reduced (DTT), Alkylated(4-VP), rpHPLC C4 3µ, TFA treated GFF support
8	ABI Procise 494 HT	0.1% TFA / 20% acetonitrile	SDS-PAGE	N/A	N/A
24	ABI Procise 494 HT	0.1%TFA/20%ACN	N/A	N/A	N/A
23	Porton 2090	0.1% TFA, 30% ACN	HPLC	None	Instrument problems
21	ABI Procise 492 HT	20% ACN, 0.1% TFA in H2O	SDS-PAGE	PGAP	Reduced, de-blocked, SDS-PAGE, blot to PVDF
38	ABI Procise 494 cLC	1x Tricine with 50 mM DTT	SDS-PAGE	None	SDS-PAGE, blot to PVDF, bands sequenced
25	ABI Procise 494 HT	25mM ABC	SDS-PAGE	None	SDS-PAGE, blot to PVDF, bands sequenced
42	Shimadzu PPSQ33-A	0.1%TFA/20% ACN	No separation	None	Spotted on PVDF, directly sequenced
43	ABI Procise 494 HT	0.1%TFA/20% ACN	No separation	None	Directly sequenced

Top-Down MS

Lab #	MS Instrument	Reconstitution Buffer	Separation	Ionization	Matrices	Data Analysis tools
PSRG-2	ABI 4800 TOF/TOF	20% ACN/0.1% TFA	HPLC	MALDI	DAN	Manual interpretation
3	ABI 4800 TOF/TOF	20% ACN/0.1% TFA, reduced 20mM DTT @60C 1hr	HPLC	MALDI	DAN	ABI 4000 Series Explorer and Manual verification
10	Bruker UltrafleXtreme	0.1 % TFA	HPLC	MALDI	SA & DAN	BioTools 3.2, T³ Mascot, P-BLAST
1	Bruker UltrafleXtreme	5% ACN 0.1% TFA	HPLC	MALDI	sDHB	Bruker Biotools and FlexAnalysis
6	Bruker Ultraflex-I	0.1%TFA	None	MALDI	SA & DAN	Bruker Biotools and FlexAnalysis

Bottom-Up MS

Lab #	MS Instrument	Analytical LC	Reconstitution Buffer	Separation	Digestion Enzymes	Data Analysis tools	Databases used
3	Thermo LTQ XL	Waters Acquity UPLC	20% ACN/0.1% TFA, reduced 20mM DTT @60C 1hr	SDS-PAGE, HPLC	Trypsin	Manual verification after performing Edman	N/A
8	Waters Q-TOF micro	Waters nanoAcquity UPLC	0.1% TFA / 20% acetonitrile	SDS-PAGE	Trypsin	N/A	N/A
21	Waters Q-TOF micro	Waters CapLC	20% ACN, 0.1% TFA in H2O	SDS-PAGE	Trypsin	MASCOT Distiller verification after performing Edman	NCBI nr and SwissProt
25	Thermo LTQ XL	N/A	25mM ABC	SDS-PAGE	Trypsin	N/A	N/A
2	Thermo LTQ Orbitrap XL	Waters nanoAcquity UPLC	Modified RIPA buffer.	SDS-PAGE	Trypsin, Pepsin, Elastase	MASCOT and Manual De Novo Sequencing.	NCBI nr
4	Bruker micrOTOF-Q II	Dionex nanoLC Ultimate 3000	25mM ammonium biocarbonate	HPLC	Trypsin	Bruker BioTools and MASCOT	N/A
PSRG-4	Thermo LTQ FT	Michrom MS4	1X Laemmli buffer	SDS-PAGE	Trypsin	Mascot	NCBI nr
15	Thermo LTQ Orbitrap XL	Waters nanoAcquity UPLC	SDS-PAGE loading buffer with DTT	SDS-PAGE	Trypsin + 6 others	Bandeira et al. Nat Biot 2008	N/A
9	Waters Synapt HDMS	Waters nanoAcquity UPLC	25mM Ammonium Bicarbonate	HPLC	Trypsin, Chymotrypsin, PNGaseF	Mascot	SwissProt
PSRG-6	Themo LTQ	Eksigent 2D-LC	98% Formic,30%peroxide	HPLC	Trypsin	Mascot	NCBI nr
PSRG-5	AB SCIEX TOF/TOF5800	N/A	SDS-PAGE loading buffer	SDS-PAGE	Trypsin	Protein Pilot and Mascot	N/A

Technique Comparison:

Edman

–Edman participants were able to directly determine the protein's N-terminal sequence. All commonly occurring amino acids including Ile/Leu and Lys/Gln as well as some stable PTMs can be identified.

–Proteins with blocked N-termini cannot be sequenced unless the protein is de-blocked. The PSRG-2010 sample's heavy chain was blocked with a pyroglutamic acid residue and required removal with pyroglutamic amino peptidase (PGAP).

–Length of sequencing read is affected by the efficiency of the instrument and size of the molecule. Read lengths were shorter than obtained by the MS techniques.

–C-terminal information cannot be obtained.

–Edman degradation does not rely on bioinformatics to obtain sequence information.

In Source Decay (top down MS)

–Matrix generated hydrogen radical mediated fragmentation of the intact protein in the ion source via laser.

–The entire ion series representing the termini may not be present and extrapolations of the ISD spectra were not sufficient to obtain the terminal sequence.

–It is necessary for ISD to use T³ sequencing to obtain the terminal sequence.

–Blocked N-termini are not an issue for ISD techniques.

–C-terminal sequence information can be obtained.

–ISD is unable to differentiate between the isobaric amino acids Ile and Leu and has difficulty differentiating between nominal mass differences found in Lys and Gln.

–Bioinformatics was relied upon to report an amino acid when in question.

Enzymatic Digestions (bottom up MS)

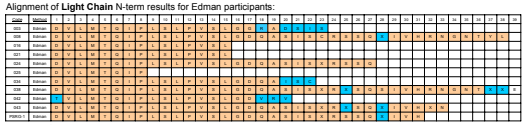
– Enzymatic or chemical cleavage of the protein followed by MS/MS analysis of the peptide mixture.
– Generates smaller fragments in random order that usually do not cover the complete protein sequence and may not include the terminal fragments.
– Successful bottom up analyses utilizes multiple enzymes and relies heavily on database homology and bioinformatics to assemble and fill in sequence gaps.
– To identify the heavy chain N-terminus through database homology, an informatic program independent of modification or the inclusion of a variable modification of an N-terminal pyro-Glutamate was necessary.
– C-terminal sequence information can be obtained.
– MS/MS is unable to differentiate between the isobaric amino acids Ile and Leu and has difficulty differentiating between nominal mass differences found in Lys and Gln unless high resolution MS/MS spectra were obtained.

Acknowledgements:

Dr. Robert English (University of Texas Medical Branch) - *Accumulation & annonimization of data*
Sigma-Aldrich - *Donation of study sample*
Executive Board - *For support and scrutiny of study proposal*
Participating labs – *For analyzing sample and returning data*

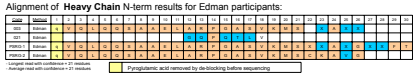
Edman Results:

Alignment of **Light Chain** N-term results for Edman participants:



Longest read with confidence = 56 residues
Average read with confidence = 24 residues

Alignment of **Heavy Chain** N-term results for Edman participants:

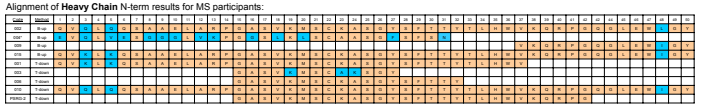


Longest read with confidence = 21 residues
Average read with confidence = 21 residues

Pyroglutamic acid removed by de-blocking before sequencing

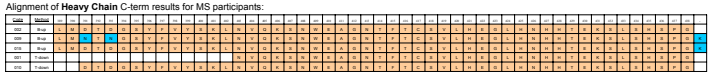
MS Heavy Chain Results:

Alignment of **Heavy Chain** N-term results for MS participants:



MS heavy chain N-terminal data reported for the first 50 residues
Participant 8 identified fragment as Ig heavy chain but was unable to find the terminal sequence
2 labs reported the presence of a C-terminal Lys
* Participant 8 reported a leader sequence not shown here for alignment purposes

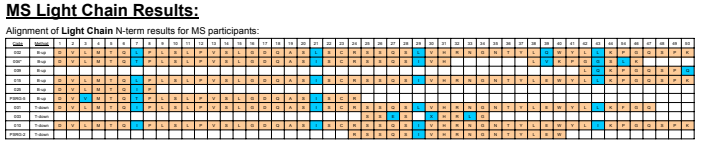
Alignment of **Heavy Chain** C-term results for MS participants:



MS heavy chain C-terminal data reported for the last 80 residues
2 labs reported the presence of a C-terminal Lys
Participant 4 was an internal sequence and did not align with the C-terminal end was not included

MS Light Chain Results:

Alignment of **Light Chain** N-term results for MS participants:



MS light chain N-terminal data reported for the first 50 residues
Participants 2 and 15 reported the full length sequence
Participant 8 identified fragment as Ig light chain but were unable to find terminal sequence
* Participant 4 reported a leader sequence not shown here for alignment purposes

Alignment of **Light Chain** C-term results for MS participants:

MS light chain C-terminal data reported for the last 80 residues
Participant 4 was an internal sequence and did not align with the C-terminal end was not included

data consistent with expected sequence | inconsistency noted | no sequence call at this position

Notes: X = multiple calls per cycle or unknown amino acid
K/Q ambiguity called as K
L/I ambiguity called as L

Conclusions:

Three techniques were successfully employed in this study to obtain terminal sequence information of an antibody not present in public databases: Edman degradation, In Source Decay (ISD, top down), and Enzymatic digestions (bottom up). All techniques used were found to be complementary.

Edman degradation analyses required de-blocking of the heavy chain using PGAP before sequencing could proceed. Some labs did not perform de-blocking due to a) cost, or b) time. Read length of Edman analyses was shorter for light and heavy chains than corresponding MS analyses, especially after heavy chain de-blocking. Edman degradation had no problem distinguishing between isobaric or nominal mass residues.

It was necessary for ISD participants to use T³ sequencing to obtain true terminal information. Extrapolation of standard ISD spectra were not sufficient to call a sequence at the termini. All MS analyses were unable to differentiate between isobaric Ile and Leu amino acids, and many had trouble identifying Lys versus Gln. Many participants relied on homology searching to determine which residue to report when in question. The most complete de novo sequencing was obtained by bottom up participants. Successful bottom up analyses utilized multiple enzymes and relied heavily on bioinformatics.

It was found that separation of the antibody chains was necessary prior to analyses to reduce the complexity of the sample to obtain both the heavy and light chain terminal sequences.

References:

1. Pham, V., Henzel, W.J., Arnott, D., Hymowitz, S., Sandoval, W.N., Truong, B., Lowman, H., and Lill, J.R. (2006). De novo proteomic sequencing of a monoclonal antibody raised against OX40 ligand. Anal. Biochem. 352, 77-86.
2. Hunkapiller, M., Hewick, R.M., Dreyer, W.J. and Hood, L.E. (1983). High-sensitivity sequencing with a gas phased sequencer. Meth. Enzymol. 91, 399-413.
3. Sukau, D., and Resemann, A. (2003). T3-sequencing: Targeted characterization of the N- and C-terminal of undigested proteins by mass spectrometry. Anal. Chem. 75, 5817-5824.