The ABRF Edman Sequencing Research Group 2009 Study:
Comparison of Edman and Mass Spectrometry Techniques for N-terminal Sequencing

RG3-S1

R. S. Thoma[1], J. S. Smith[2], W. Sandoval[3], B.S. Hampton[4], N. D. Denslow[5], P. Hunziker[6]

[1]Monsanto Co., St. Louis, MO, United States, [2]University of Texas Medical Branch, Galveston, TX, United States, [3]Genentech Inc., South San Francisco, CA, United States, [4]University of Maryland School of Medicine, Baltimore, MD, United States, [5]University of Florida, Gainesville, FL, United States, [6]University of Zurich, Zurich, Switzerland.

Introduction

For decades, high sensitivity Edman sequencing has been the method of choice for determining the N-terminal amino acid sequence of proteins[1]. However, the major but by no means only limitation of this technique is its inability to obtain amino acid sequence from N-terminally blocked proteins. Mass spectrometric techniques for protein sequence analysis do not suffer from this limitation though unequivocal determination of protein N-termini on a routine basis has been elusive. The advantages of mass spectrometric techniques have in recent years driven investigators to look beyond Edman chemistry to find alternative technologies to obtain N-terminal sequence. Several mass spectrometric methodologies have been published, primarily for proteomics analyses, which may be quicker, less costly and more sensitive than Edman sequencing. Because such techniques involve a range of biochemical and instrumental methodologies having different advantages and limitations the ESRG has created a study to ascertain how reliably they can produce N-terminal amino acid sequence information and to compare those results to those obtained by automated Edman sequencing.

The ESRG 2009 study was designed to allow the participants freedom to use their analytical technique of choice to obtain as much N-terminal amino acid sequence information as possible from two test proteins in solution. Approximately one nanomole of each sample was provided so laboratories were able to attempt a variety of techniques with the goal of obtaining each protein's N-terminal sequence. Results of the analyses from participating laboratories are illustrated. Included is a comparison of methodology and instrumentation used by the laboratories to determine the N-terminal sequence of the protein.

Materials and Methods

TEST SAMPLE 1

E.coli BL21 was transformed with a pTrcHisTOPO plasmid containing a cDNA encoding the yeast alcohol dehydrogenase (ADH1) protein. Expression resulted in an ADH1 protein with a 35 amino acid leader sequence containing a hexahis-tag.

Alcohol Dehydrogenase 1 (Yeast) SP_P00330

GGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPTLMSIPETQKGVIFYESHGKLEYKDIPVPK
PKANELLINVKYSGVCHTDLHAWHGDWPLPVKLPLVGGHEGAGVVVGMGENVKGWKIGDYAG
IKWLNGSCMACEYCELGNESNCPHADLSGYTHDGSFQQYATADAVQAAHIPQGTDLAQVAPI
LCAGITVYKALKSANLMAGHWVAISGAAGGLGSLAVQYAKAMGYRVLGIDGGEGKEELFRSI
GGEVFIDFTKEKDIVGAVLKATDGGAHGVINVSVSEAAIEASTRYVRANGTTVLVGMPAGAK
CCSDVFNQVVKSISIVGSYVGNRADTREALDFFARGLVKSPIKVVGLSTLPEIYEKMEKGQI
VGRYVVDTSK

BOLD = 6-histidine-tagged beta galactosidase (cloning vector)

After transformation, *E. coli* was incubated in the presence of ampicillin (50µg/mL) and expression was induced with IPTG (0.5mM) in a New Brunswick Bioflow III fermentor with a 5L vessel. *E coli* cells were harvested by centrifugation. Cells were re-suspended and lysed in a 25mM Tris-HCl pH 8.0, 0.3 M NaCl, 10mM Imidazole, 0.5% Triton-X 100. The lysate was loaded on a Ni Sepharose High Performance (GE Healthcare) resin column. The Ni column was washed with 25mM Tris-HCl pH 8.0, 0.3 M NaCl, 10mM Imidazole. After washing, the ADH1 protein was eluted with 25mM Tris-HCl pH 8.0, 0.3 M NaCl, 250mM Imidazole. Triton-X 100 was removed by washing 3X with Bio-Beads SM-2 Adsorbent (BioRad). ADH1 purity was assessed with MALDI-TOF, SDS-PAGE gel and Edman sequencing. The ADH1 was quantitated by amino acid analysis. One nmol of ADH1 was aliquoted in 1.5 ml low protein retention microcentrifuge tubes (Fisher) and dried for distribution to the study participants.

TEST SAMPLE 2

Glyceraldehyde-3-phosphate dehydrogenase, Rabbit Muscle, (G3P) was purchased from Sigma. (Cat. No. G5262)

Glyceraldehyde-3-phosphate dehydrogenase (Rabbit) SP_P46406

VKVGVNGFGRIGRLVTRAAFNSGKVDVVAINDPFIDLHYMVYMFQYDSTHGKFHGTVKAENG
KLVINGKAITIFQERDPANIKWGDAGAEYVVESTGVFTTMEKAGAHLKGGAKRVIISAPSAD
APMFVMGVNHEKYDNSLKIVSNASCTTNCLAPLAKVIHDHFGIVEGLMTTVHAITATQKTVD
GPSGKLWRDGRGAAQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTPNVSVVDLTCRLEK
AAKYDDIKKVVKQASEGPLKGILGYTEDQVVSCDFNSATHSSTFDAGAGIALNDHFVKLISW
YDNEFGYSNRVVDLMVHMASKE

G3P was weighed and re-suspended in 0.1% TFA. One nmol was aliquoted in 1.5 mL low protein retention microcentrifuge tubes (Fisher) and dried for distribution to the study participants.

Table 1. Sequence Call by Participating Laboratory

a. Test Sample 1 – Alcohol Dehydrogenase (Yeast) with a 6 Histidine-tag β-Galactosidase (cloning vector)

Analysis Method	ESRG Lab #	N-terminal Sequence Data: G G S H H H H H H G M A S M T G G Q Q M G R D L Y D D D D K D P T L M S I P E T Q K G V I F Y E S H G K L E Y K D I P V > 60	Correct Protein ID	# Correct AA Calls
Bottom-Up*	001		Yes	13
	005		Yes	12
	011		Yes	7
	016		Yes	9
	018		Yes	11
	C			
Edman	006		No	30
	008		No	34
	009		No	34
	014		No	42
	016		No	18
	017		No	29
	018		No	25
	020		No	22
	023		No	0
	025		No	31
	026		No	30
	A		No	30
	B		No	34
	C		No	30
	D		No	21
Top-Down**	001		Yes	44
	002		Yes	50
	004		Yes	19
	005		Yes	16
	007		Yes	15
	008		No	18
	010		Yes	18
	012		Yes	21
	013		Yes	0
	015		Yes	16
	019		Yes	18
	021***		Yes	18
	022		Yes	31
	024		Yes	17
	C		No	9

b. Test Sample 2 – Glyceraldehyde 3-Phosphate Dehydrogenase (Rabbit)

Analysis Method	ESRG Lab #	N-terminal Sequence (Data): V K V G V N G F G R I G R L V T R A A F N S G K V D V V A I N D P F I D L H Y M V Y M F Q Y D S T H G K F H G T V K A E >60	Correct Protein ID	# Correct AA Calls
Bottom-Up*	001		Yes	8
	005		Yes	9
	010		Yes	9
	011		Yes	10
	016		Yes	8
	C		Yes	9
Edman	006		Yes	30
	008		Yes	35
	009		Yes	49
	014		Yes	78
	016		Yes	30
	017		Yes	41
	018		Yes	33
	020		Yes	24
	023		Yes	26
	025		Yes	40
	026		Yes	26
	A		Yes	64
	B		Yes	24
	C		Yes	37
	D		Yes	18
Top-Down**	001		Yes	50
	002		Yes	54
	004		Yes	28
	005		Yes	32
	007		Yes	17
	008		Yes	9
	010		Yes	30
	012		Yes	26
	013		Yes	22
	015		Yes	32
	019		Yes	25
	021***		Yes	26
	022		Yes	28
	024		Yes	12
	C		Yes	13

Box color code: green = Correct amino acid call based on sequence data alone; blue = Protein ID from a data base search was used to identify the amino acid; red = Incorrect amino acid call

Labs A, B, C and D data were contributed by ESRG committee members

* ESRG -003, -023, -025 and -026 Protein identified from Bottom-Up data (not shown)

** Many laboratories using the Top-Down technique also supplied C-terminal sequence data (data not shown)

*** ESRG-021 evidence for the complete N-terminal sequence not present in the MS/MS spectra.

Table 2. Instrumentation and Other Parameters used by Participating Laboratories

Analysis Method	ESRG Lab #	Instrumentation and Methods								
Bottom-Up		**Manufacturer**	**Instrument**	**Ionization**	**Reflector/linear**	**Digestion Enzymes**	**Digestion Solvent**	**Red. / Alkyl.**	**Ions measured**	
	001	Waters-Micromass	QTOF-2 Nano	ESI		Trypsin	400mMNH_4HCO_3	none	b & y	
	003	Waters-Micromass	Q-TOF Premier	ESI		Trypsin & Glu-C	0.4MNH_4HCO_3/8M Urea	DTT & IAM		
	005	Bruker	Ultraflex-II	MALDI	reflector	Trypsin			b & y	
	011	Thermo-Electron	LTQ-FT	ESI		Chymotrypsin			b & y	
	016	Thermo-Fisher	LTQ-Orbitrap	ESI		Trypsin, Chymo. & Glu-C			b & y	
	018	Waters-Micromass	Q-TOF-micro	ESI		Trypsin			a,b y, & z	
	023	Thermo-Finnigan	LCQ-Duo	ESI		Trypsin				
	026	Waters-Micromass	Q-TOF-micro	ESI		Trypsin	20%ACN/20mMTris, pH8			
	C	Thermo-Fisher	LTQ-Orbitrap	ESI		Trypsin	0.1M Tris	none		
Edman		**Manufacturer**	**Instrument**	**Column**	**Chemistry Cycle**	**TFA Deleivery**	**Sample Solvent**	**Sample Support**		
	006	ABI	491-HT	Spheri-5 PTH	PVDF peptide	liquid	0.1%TFA	PVDF		
	008	ABI	494-HT	Spheri-5 PTH	pulsed liquid PVDF	liquid	0.1%TFA	PVDF		
	009	ABI	494-HT	Spheri-5 PTH	na	liquid	0.1%TFA	TFA treated GFF		
	014	ABI	494-cLC	Procise cLc	PVDF peptide	liquid	70% Formic acid	PVDF		
	016	ABI	494-HT	Higgins HL-C18	PVDF peptide	gas + liquid	na	PVDF		
	017	ABI	492-cLC	Procise cLc	PVDF	na	0.2%TFA	Biobrene Treated GFF		
	018	ABI	494-HT	Spheri-5 PTH	PVDF + GFF	liquid	0.1%TFA	1-Prosorb/2-GFF		
	020	ABI	494-HT	Spheri-5 PTH	GFF	pulsed liquid	0.1%TFA	GFF		
	023	ABI	492-HT	Spheri-5 PTH	GFF	gas	0.1%TFA	GFF		
	025	ABI	494-cLC & 492-HT	cLC/Spheri-5 PTH	pulsed liquid PVDF	liquid	na	PVDF		
	026	ABI	494-HT	Spheri-5 PTH	GFF liquid	liquid	20%ACN/0.1%TFA	GFF		
	A	ABI	494-HT	Spheri-5 PTH	PVDF peptide	liquid	0.1%TFA	Biobrene Treated Prosorb		
	B	ABI	494-HT	Spheri-5 PTH	pulsed liquid PVDF	liquid	20%ACN/0.1%TFA	GFF		
	C	ABI	494-HT	Higgins 15cm	pulsed liquid PVDF	liquid	20%ACN/0.1%TFA	PVDF		
	D	ABI	492-cLC	Procise cLC	pulsed liquid PVDF	liquid	20%ACN/0.1%TFA	PVDF-Prosorb		
Top-Down		**Manufacturer**	**Instrument**	**Ionization**	**Analyzer**	**Reflector/linear**	**Solubilization Buffer**	**Matrix**	**N-term**	**C-term**
	001	Bruker	Ultraflex-II	MALDI	TOF-TOF	reflector ISD	0.1%TFA	DHB	c	na
	002	Bruker	Ultraflex-III	MALDI	TOF-TOF	reflector ISD	50%ACN/0.1%TFA	DHB	a & c	y & z+2
	004	Bruker	Apexultra 7T	MALDI	FTICR	reflector ISD	0.1%TFA	DAN	c	y & z+2
	005	Bruker	Ultraflex-II	MALDI	TOF-TOF	reflector ISD	na	DHB	c	na
	007	Bruker	Ultraflex-II	MALDI	TOF-TOF	linear & reflector ISD	0.1%TFA	DHB & DAN	c	z
	008	Bruker	Ultraflex-II	MALDI	TOF-TOF	reflector ISD	50%ACN/0.1%TFA	DHB & DAN	a & c	y & z
	010	Bruker	Ultraflex-III	MALDI	TOF-TOF	reflector ISD	2%ACN/0.1%TFA	SA	c	y
	012	Bruker	Ultraflex	MALDI	TOF-TOF	reflector ISD	50%ACN/0.1%TFA	DHB	c	y
	013	Bruker	Ultraflex-II	MALDI	TOF-TOF	reflector ISD	0.1%TFA	DHB	a & c	y & z+2
	015	Bruker	Microflex	MALDI	TOF	reflector ISD	50%ACN/0.1%TFA	DHB &SA	c	y & z+2
	019	Bruker	Autoflex	MALDI	TOF-TOF	reflector ISD	10%ACN/0.1%TFA	DHB	c	y
	021*	Thermo-Fisher	Orbitrap	ESI	Linear ion-trap	ETD	50%ACN/0.1%Formic acid	infusion	c	na
	022	Bruker	Ultraflex-III	MALDI	TOF-TOF	reflector ISD	na	Super DHB	a, b & c	y & z+2
	024	Bruker	Ultraflex-I	MALDI	TOF	linear & reflector ISD	0.1%TFA	DHB	c	na
	C	ABI	4800	MALDI	TOF-TOF	reflector ISD	50%ACN/0.1%TFA	DAN & PA	c & z	na

Fig. 1 Example Data Supplied for the Edman Degradation Chemistry Approach from ESRG-A – Test Sample 2

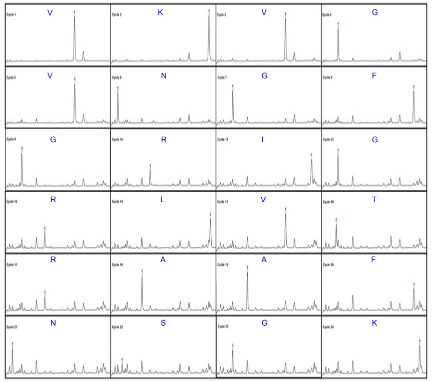
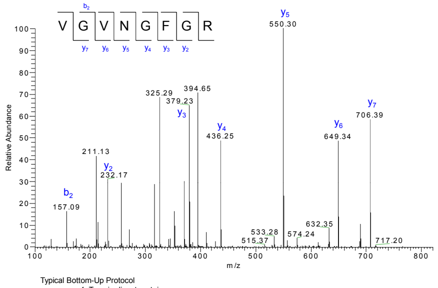


Fig. 2 Example Data Supplied for the Bottom-Up Approach from ESRG-016 – Test Sample 2



Typical Bottom-Up Protocol
1. Trypsin digest protein
2. LC-MS/MS
3. Identify protein from the mass fingerprint
4. Validate presence of the N-terminal peptide by manual or automatic inspection of the data.

Fig. 3 Example ISD Data Supplied for the Top-Down Approach from ESRG-002 – Test Sample 1

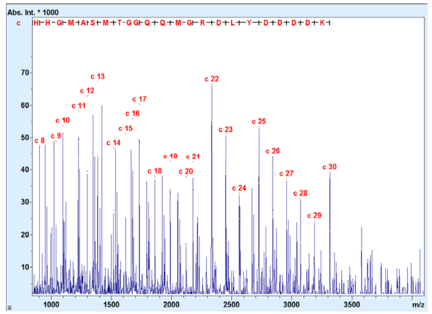


Table 3 Comparison of the Techniques used to Determine the N-terminal Sequence

Sequence Analysis Technique	Mechanism	Advantages	Disadvantages
Edman Degradation[1]	Sequential chemical degradation of the protein by coupling of PITC to the N-terminus followed by cleavage with TFA and identification of the resulting phenylthiohydantoin-amino acid by liquid chromatography and UV detection.	Direct determination of the protein N-terminal sequence. All commonly occurring amino acids identifiable. Stable PTMs identified by retention time compared to the known modified PTH-AA standard.	N-terminal blockage (e.g. N-acetylation) prevents sequencing. Length of sequence read limited by coupling/cleavage efficiency. Certain PTMs problematic. Requires high purity chemicals - high cost. Low throughput - 20 to 40 minutes per cycle for each residue. Commercial production of instrument ceased June 2008.
Top-Down MALDI In-Source Decay[2,3]	Matrix generated hydrogen radical mediated fragmentation of intact protein in the ion source typically with higher laser power and chemical additives. Fragmentation occurs in the nanosecond time scale.	Rapid, high throughput, low cost. C-terminal sequence can also be determined. Stable PTMs identifiable. Isolation of terminal fragments for ms/ms possible but is instrument configuration dependant. Instruments commonly available.	ISD doesn't provide precursor ion selection - sample must be purified. May not yield complete N-terminal sequence i.e. protein terminii must be accessible to ionization/fragmentation process. Cannot distinguish isobaric amino acids. Background ion interference in the low mass region - the entire ion series representing the N-terminus may not be present.
Top-Down ESI Electron Transfer Dissociation[4]	Electon transfer from Anthracene anions to multiply protonated precursor cations resulting in exclusive peptide backbone fragmentation. Fragmentation occurs in the millisecond time scale.	Preserves PTM structures, rapid, high throughput. Precursor ion selection allows MS^n of known precursor ion mass. Extensive peptide backbone fragmentation.	ETD fragmentation requires multiply protonated precursors. For MS^n of CAD generated lower charge state precursors ETD efficiency decreases - the entire ion series representing the N-terminus may not be present. Cannot distinguish isobaric amino acids. Relatively new technique - fewer instruments in use. Requires high resolution mass analyzer to obtain charge state of protein precursor - higher cost of instrumentation. Only structurally accessible portions of the precursor are fragmented.
Bottom-Up ESI MS/MS MALDI TOF/TOF MALDI PSD[5]	Enzymatic or chemical cleavage of the protein followed by ms/ms analysis of the peptide mixture and bioinformatic analysis to identify the N-terminus. Biochemical techniques to obtain the N-terminal peptide prior to ms/ms analysis.	Generates smaller fragments that are amenable to analysis on a wide variety of instrumentation. High throughput, low cost depending on specific technique used.	Successful isolation of the N-terminal peptide is sequence dependent, possibly requiring multiple biochemical approaches. Isobaric amino acids not determined. Absence of the complete ion series in the ms/ms spectrum is common and prevents interpretation of the amino acid sequence for the entire peptide.

Conclusions

1. Edman sequencing remains a reliable means for determining the N-terminal sequence of an unblocked protein. The majority of labs that used this approach were able to easily sequence both test proteins. However, no one that used this technique was able to identify Test Sample 1 because of the long vector sequence.
2. The Top-Down approach shows great promise for determining the N- (and C) terminal sequence of a protein in solution. The majority of labs using this approach correctly identified long strings of amino acids near the N- (and C) terminus. Low mass ions (< 700 m/z) were poorly resolved, making the first 5-7 amino acids difficult to call. Most labs easily identified both proteins.
3. The Bottom-Up approach (without modification to the N-terminal amino acid) works well when the mass fingerprint can be matched to a protein in a data base. Those labs using the bottom-up approach found the N-terminal fragment of Test Sample 2 and were able to MS/MS sequence this fragment.
4. Labs using a mass spectrometry approach often depend on the protein ID from a data base search to fill in the following:
   a. Amino acids with isobaric masses (Ile/Leu and Gln/Lys)
   b. Regions with poor ion signal (peptide bonds associated with Pro)
   c. Missing low mass ions (< 700 m/z for ISD)
5. For those performing Edman chemistry, sequence length was typically shorter for Test Sample 1 (contained a 6 histidine-tag) than Test Sample 2. This matches observations from the ESRG 2008 study that found histidine-tags are more difficult to sequence than other proteins. The histidine-tag was not a problem for those using an MS approach.

References

1. Hunkapillar, 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.
2. Reiber, D.C., Grover, T.A., Brown, R.S. (1998). Identifying proteins using matrix-assisted laser desorption/ionization in-source fragmentation data combined with database searching. Anal. Chem. 70, 673-683.
3. Suckau, D., and Resemann, A. (2003). T3-sequencing: Targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. Anal. Chem. 75, 5817-5824.
4. Syka et al. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci.101 (26), 9528-9533.
5. Yamaguchi et al. (2008). Selective isolation of N-terminal peptides from proteins and their de novo sequencing by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry without regard to unblocking or blocking of N-terminal amino acids. Rap. Com. Mass Spec. 22 (20), 3313-3319.

Acknowledgements

The ESRG would like to express our gratitude to the following people:
Bo Xu (University of Texas Medical Branch) and **PERG** for Test Sample 1 production
Robert English (University of Texas Medical Branch) for data accumulation
All **Participating labs** for analyzing samples and returning data
The **ABRF Executive Board** for support and scrutiny of study proposal