The ABRF Edman Sequencing Research Group 2008 Study: Investigation into Homopolymeric Amino Acid N-terminal Sequence Tags and their Effects on Automated Edman Degradation

J. S. Smith[1], R. S. Thoma[2], W. Sandoval[3], J. W. Leone[4], K. D. Linse[5], P. Hunziker[6], N. D. Denslow[7], B. Hampton [8]

[1]University of Texas Medical Branch, Galveston, TX, United States, [2]Monsanto Co., St. Louis, MO, United States, [3]Genentech Inc., South San Francisco, CA, United States, [4]Pfizer Inc., St. Louis, MO, United States, [5]University of Texas, Austin, TX, United States, [6]University of Zurich, Zurich, Switzerland, [7]University of Florida, Gainesville, FL, United States, [8]University of Maryland School of Medicine, Baltimore, MD, United States.

Introduction

For decades, Edman degradation(1) has been an invaluable tool for protein characterization. Though other techniques have surpassed Edman chemistry in ease, cost, and utility for routine protein characterization, automated Edman degradation remains the most effective tool for obtaining N-terminal amino acid sequence information. A common affinity tag for protein purification is a poly-Histidine sequence which may be conjugated to either termini of the protein. After expression, His-tagged proteins are readily purified via chelation with an immobilized metal affinity resin. Determining the N-terminal sequence for several amino acids beyond the His tag is important for confirming the proper expression of the protein. N-terminal His tags are often found to be problematic with regards to Edman sequencing with poor repetitive yields and overall low signal. Problems with sequencing preview have also been reported when sequencing Histidine containing proteins.(2) Therefore, the question is posed: is this the result of a homopolymeric amino acid sequence in general, or specifically the presence of histidine in the affinity tag sequence? The Edman Sequencing Research Group (ESRG) of ABRF has enlisted the help of core sequencing facilities to investigate the effects of a repeating amino acid tag at the N-terminus of a protein. The laboratories were asked to sequence the same protein engineered in three configurations: 1) with an N-terminal poly-His tag, 2) an N-terminal poly-Ala tag, or 3) no tag. Study participants were asked to return a data file containing the uncorrected amino acid picomole yields for the first seventeen cycles. Initial and repetitive yield information and the amount of lag were evaluated. Information on instrumentation and sample treatment were also collected.

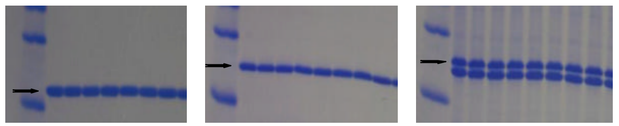
Materials and Methods

Cloning, Expression and Purification:
Homo-poly amino acid DNA constructs were created in house. The following Human Growth Hormone (hGH) PCR primers were used:

hGH-N-Term. His-Tag.XhoI>CTCGAGTTCCCAACCATTCCCTTATCC
hGH-polyAla-Tag.ClaI.F>CCATCGATGCTGCAGCTGCAGCTGCAGCTGCATTCCCAACCATTCCCTTATCC
hGH-polyLys-Tag.ClaI.F>CCATCGATAAGAAGAAGAAGAAGAAGAAGAAGTTCCCAACCATTCCCTTATCC
hGH-polyTyr -Tag.ClaI.F>CCATCGATTACTACTACTACTACTACTACTACTACTTCCCAACCATTCCCTTATCC

PCR reaction was performed with Clontech Advantage GC polymerase mix. After PCR, 5µL of the reaction was run on a gel to visualize the product. DNA was further purified using a PCR cleanup kit. The DNA was subjected to restriction digests, as were the pRK.sm vectors (1.5h at 37ºC). Products were ligated at a 1:3 ration of vector to insert (total volume 10µL), and incubated overnight at 14°C. The DNA was then transfected into HEK 293 cells using Qaigen Polyfect. For each 150cm^2 of cells, 0.6mL of serum free 50:50 media containing 16µg of DNA was mixed with 160µL Polyfect and incubated at RT for 10 minutes. 10ml of fresh complete media was added to each plate during the incubation. After incubation, 1ml complete media was added dropwise onto cells. Following a three day incubation, the media which was removed from the plates was incubated with washed Ni-NTA resin (4°C, 2h). Solutions were spun, washed with PBS and the protein was eluted with 250mM imidazole in PBS.

Control "Sample C" | **N-term His tag "Sample H"** | **N-term Ala tag "Sample A"** (top band)



Note: Arrows indicate bands excised and sent to study participants. Top band from sample A sent to participants. "Sample A" bottom band was found to have the same N-terminal sequence as the top band.

Processing and distribution:
Test proteins were analyzed by SDS-PAGE stained with Coomassie Brilliant Blue. Approximate concentrations were determined based on the intensity of the Control Protein. Samples were reduced (10mM DTT), and alkylated (0.2M N-isopropyl iodacetamide) in sample buffer. 25pmol of the control, along with the visual equivalent of 25 pmol for His-tagged and Ala-tagged samples, were loaded onto multiple gels (4-20% tris-glycine) and electroblotted onto PVDF. Two bands were excised from each blot and sent to participating laboratories.

Sequences:
Sample H: K-H-H-H-H-H-H-H-H-L-E-F-P-T-I-P-L
Sample A: K-I-D-A-A-A-A-A-A-A-A-F-P-T-I-P-L
Sample C: F-P-T-I-P-L-S-R-L-F-D-N-A-M-L-R-A

Sequence Call by Facility

Lag not evident | Lag overtaken yield

	70	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	p	L	492 cLC	liquid	1
	100	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I		L	492 cLC	liquid	1
	ESRG2	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	P	L	492 cLC	liquid	1
	ESRG7	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	P	L	492 cLC	liquid	1
note 3	300	K	h*	h*	h*	h*	h*	h*	h*	h*	L	E	F	P	T	I	P	L	494 cLC	liquid	1
	400	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	p	L	494 cLC	liquid	1
	10	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	d	n	494 HT	liquid	2
	30	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	i	r	494 HT	liquid	2
	40	K		H	H	H	h	h/	h		L	E	F	P	I	P	L	S	491 HT	liquid	2
	60	K	H	H	H	H	H	H	H	H	L	E	F	P	T	T	P	L	494 HT	gas	1
	80	K	H	H	H	H	H	H	H	H	L	E	F	P	T	T	P	L	494 HT	gas	2
	90	K	H	H	H	H	H	H	H	h	L	E	F	P	T	I	P	L	494 HT	liquid	2
note 2	200	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	p	L	494 HT	liquid	2
	500	K	H	H	H	H	H	H	H	H	L	E	F	P	K		K	K	494 HT	liquid	2
note 7	600	L	H	H	H	H	H	H	H	H	L	E	F	P	T	I		L	494 HT	liquid	2
note 4	ESRG1	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I	P	L	494 HT	gas	2
	ESRG4	K	H	H	H	H	H	h	H	H	L	E	F	P	T	I	p	L	494 HT	liquid	2
	ESRG5	K	H	H	H	H	H	H	H	H	L	E	F	P	t				494 HT	liquid	2
	ESRG6	K	H	H	H	H	H	H	H	H	L	E	F	P	T	I		L	494 HT	liquid	1
note 5	ESRG3	K	H	H	H	H	H	H	h	h	L	E	F	P	T	I			494 HT	liquid	2

lag not evident
lag overtaken yield
(* -co-eluted with sLYS standard)
(lower case = tentative call)

Sample A	Cycle ->	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Instrument	TFA	Bands
Expected sequence		K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L		Delivery	loaded
	Facility																				
note 1	20	K	I	D	A	A	A	A	A	A	L	A	F	P	T	I	P	L	492 cLC	liquid	2
	50	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	492 cLC	liquid	1
	70	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	492 cLC	liquid	1
	100	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	492 cLC	liquid	1
	ESRG2	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	492 cLC	liquid	1
	ESRG7	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	492 cLC	liquid	1
note 3	300	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 cLC	liquid	1
	400	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 cLC	liquid	1
	10	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
	30	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
	40	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	491 HT	liquid	2
	60	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	gas	1
	80	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	gas	2
	90	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
note 2	200	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
	500	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
	600	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	F	L	494 HT	liquid	2
note 4	ESRG1	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	gas	2
	ESRG3	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
	ESRG4	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
	ESRG5	K	I	D	A	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	2
note 5	ESRG6	K	I	D	A	A	A	A	A	A	A	A	F	P	T	I	P	L	494 HT	liquid	1

Sample C	Cycle ->	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Instrument	TFA	Bands
Expected sequence		F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A		Delivery	loaded
	Facility																				
note 1	20	F	P	T	I	P	L	S	R	L	F	D	N	A	m/I	I/r	r/a	a/r	492 cLC	liquid	2
	50	F	P	T	I	P	L	S	r	L	F	D	N	A	M	L	r	A	492 cLC	liquid	1
	70	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	492 cLC	liquid	1
	100	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	492 cLC	liquid	1
	ESRG2	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	492 cLC	liquid	1
	ESRG7	no data returned																			
note 3	300	F	P	T	I	P	L	S		L	F	D	N	A	M	L		A	494 cLC	liquid	1
	400	F	P	T	I	P	L	S	r	L	F	D	N	A	M	L	L	A	494 cLC	liquid	1
	10	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	liquid	2
	30	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	F	494 HT	liquid	2
	40	F	P	T	I	P	L	S		L	F	D	N	A	M	L	R	A	491 HT	liquid	2
	60	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	gas	1
	80	F	P	T	I	P	L	S	r	L	F	D	N	A	M	L	r	A	494 HT	gas	2
	90	F	P	T	I	P	L	S	r	L	F	D	N	A	M	L	r	A	494 HT	liquid	2
note 2	200	no data returned																			
	500	no data returned																			
note 6	600	T	I	P	L	S															
note 4	ESRG1	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	gas	2
	ESRG3	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	liquid	2
	ESRG4	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	liquid	2
	ESRG5	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	liquid	2
note 5	ESRG6	F	P	T	I	P	L	S	R	L	F	D	N	A	M	L	R	A	494 HT	liquid	1

Table 1 Sequence calls by facility. Residues shaded red indicate cycles where the sequencing lag has overtaken the yield.The majority of participating labs successfully called the amino acid sequence for seventeen cycles for all three test proteins

(lower case = tentative call; blank space = no call); (* -co-eluted with succinylated LYS from added peptide standard)
note 1: Samples were de-stained 3 x 1 min. with 100% Methanol
note 2: Direct load 2 bands into std. Cart. With Zitex seal only
note 3: PVDF samples slit and loaded above a pre-treated glass fiber filter
note 4: Samples were sequenced in the blot cartridge
note 5: Extended R1 coupling in first amino acid cycle after begin cycle in all samples. Five R1 deliveries (three R1 deliveries in normal cycle) and doubled the coupling time to 340 sec (170 sec coupling time in normal cycle)
note 6: Sequencer reported to have R3 delivery problems
note 7: Could not calculate cycle lag. Many zeros in raw data

Percent Lag at Phe

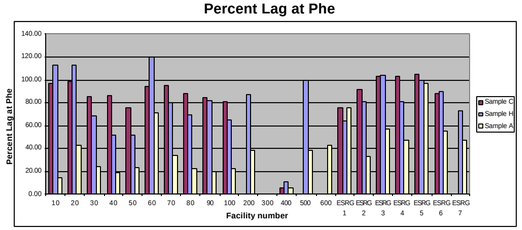


Figure 1: Percentage of lag for the His-tagged protein, Ala-tagged protein, and control protein. Calculation of percent lag is based on amount of lag present in the cycle following the Phe at cycle 12 for the His-tagged and Ala-tagged samples, and cycle 10 for the control sample.

Ala-Tag vs. His-Tag Sequencing Yield

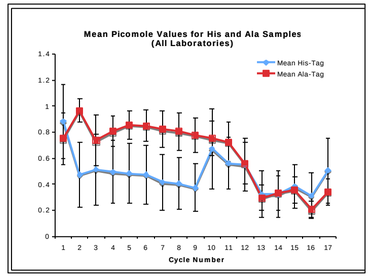


Figure 2 Normalized mean picomole values for the poly Ala-tagged sample vs. the poly His-tagged sample. A noticeable decrease in picomole yield from cycle 1 to cycle 2 was observed in the His-tagged sample. This decrease in yield was not present in the Ala-tagged sample. Error bars represent sample variation between facilities.

Average Repetitive Yields

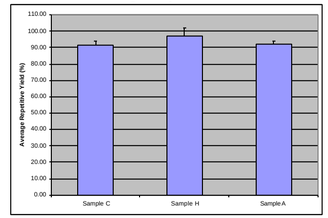


Figure 3: Average repetitive yields for Sample C, Sample H, and Sample A. Higher repetitive yield for Sample H is a result of the slight preview observed in this set of samples. (see Figure 5)

Average Initial Yield

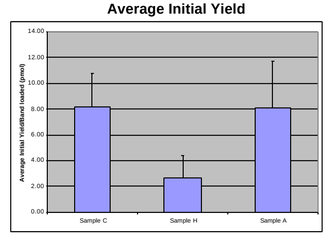


Figure 4: Average initial yields for Sample C, Sample H, and Sample A. The lower initial yield observed with Sample H contributed to the difficulty experienced by some facilities in calling the sequence after the poly His-tag.

Yield of Phenylalanine in All Cycles

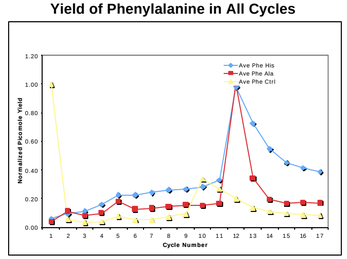


Figure 5: Normalized picomole yield of Phe in all cycles. A slight preview, as well as significant lag is seen in the His-Tag sample verses the control samples. These factors contributed to the difficulty experienced by some facilities in calling the sequence after the poly His-tag.

Instrument Statistics

Manufacturer and Model	13	ABI 494 HT (2.5 - 15 years old: average age 10.4 years)
	6	ABI 492 cLC (5-10 years old: average age 6.5 years)
	2	ABI 494 cLC (8 and 13 years old: average age 10.5 years)
	1	ABI 491 HT (5 years old)
Reagents	18	used all instrument manufacturer reagents
	3	used some manufacturer reagents
	1	said R2, R4, S4 home made
TFA Cleavage	19	used pulsed liquid
	3	used gas phase
Chemistry Cycle	16	PVDF
	3	GFF
	1	PVDF & Prosorb
	1	PVDF & Prosorb
Other Additives	1	TCEP to R4 and R5
	1	DTT in S2B/ethyl acetate
	1	n-acetylcysteine in R5
	1	34 % N-Methyl-Piperidine, 52 % isopropanol in R2
Columns Used	11	ABI Spherisorb 5 micron PTH column (2.1 x 220 mm)
	1	Higgins column 3 micron (2.1 x 100)
	1	Higgins column 5 micron (2.1 x 220)
	1	Higgins column 5 micron (2.1 x 250)
	8	ABI Prosice cLC PTH column 5 micron (0.8 x 250)

Table 2 Data extracted from surveys returned by participating labs. Labs either used ABI HT or ABI cLC sequencers. Most used the manufacture's reagents and pulsed-liquid PVDF chemistry cycles. Only four facilities reported using additives and most used the manufactures PTH column.

Typical Gradients Used

HPLC Gradient HT

Time	%B	Flow Rate (ul/min)	
0	8	325	
0.3	8	325	
0.4	23	325	
18	44	325	
22	95	325	
22.5	95	325	
26	50	10	

HPLC Gradient cLC

Time	%B	Flow Rate (ul/min)	
0	10	40	
0.4	12	40	
4	22	40	
22	46	40	
22.6	90	40	
23.5	90	40	
29	90	40	
33	50	5	

Conclusions

•Creation of a poly-amino acid-tagged recombinant protein is not easy to do. Attempts to express poly-Lysine and poly-Tyrosine tagged proteins were not successful. Other than the traditional His-tag, the only other successfully prepared poly-amino acid tag for this study was an Ala-tag. (data shown in presentation).
•The majority of participating labs successfully called the amino acid sequence for seventeen cycles for all three test proteins. (Table 1)
•Labs, in general, found it harder to call the sequence after the poly-His-tag than the other two test samples.
•Lag was observed earlier and more consistently on the poly-His-tag protein than the poly-Ala-tag protein. (Table 1, Figure 1)
•There was a noticeable decrease in yield from cycle 1 to cycle 2 of the poly-His-tag protein. This decrease was not observed in the poly-Ala-tagged protein. (Figure 2)
•Averaged normalized Phenylalanine values from each cycle indicated a significant increase in lag with the His-tagged sample as compared to the other two test samples. (Figure 5)
•There was evidence of a low level sequencing preview of Phe in the His-tag sample. (Figure 5)
•Large error bars in the figures are a reflection of high variability in amino acid yields between participating labs. (Figure 2, Figure 3, Figure 4)
•No significant correlation was observed between the appearance of sequence lag with either the type of instrument used, age of instrument, reagent additives or chemistry used.(Table 1, Table 2)

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. Thomsen, J., Kristiansen, K., Brunfeldt, K. (1972) The Amino Acid Sequence of Human Glucagon FEBS Letters **21**, 315-319.

Acknowledgments

Jessica Huard - Genentech- Expression and Purification of study samples
Liza Ingle - Genentech - Running gels and blots for distribution
Glenda Cowart - Monsanto - Accumulation of data
Participating labs - Analyzing samples and returning data
Executive Board - For support and scrutiny of study proposal
ESRG members - Design and execution of study