# The ABRF Edman Sequencing Research Group 2008 Study

## Effects of a Homopolymeric amino acid tag on Edman Degradation

## Current ESRG Members

- Richard S. Thoma (Co-Chair) — *Monsanto*
- Brian Hampton (Co-Chair) — *University of Maryland School of Medicine*
- Joseph W. Leone (Chair-emeritus) — *Pfizer, Inc.*
- Peter Hunziker — *University of Zurich*
- Klaus Linse — *University of Texas - Austin*
- Wendy Sandoval — *Genentech, Inc.*
- J. Steve Smith — *University of Texas Medical Branch - Galveston*
- Nancy D. Denslow (EB liaison) — *University of Florida - Gainesville*

# What is the ESRG?

- **Mission of the Edman Sequencing Research Group:**

  "… to assist ABRF members in evaluating their capabilities to analyze the N-terminus of proteins/peptides using Edman Sequencing chemistry in order to establish realistic expectations for this technology. "

- **ESRG primary responsibility: a multi-lab participation educational study**
  - Sample Conception
  - Study Design (approval by EB)
  - Sample Preparation and distribution
  - Preliminary testing by group
  - Data collection, analysis and interpretation
  - Data presentation
    - Poster Presentation at ABRF meeting (Poster # **RG8-M**)
    - Oral Presentation at ABRF meeting
    - Publication in JBT (upcoming)

# 20 ESRG Studies: 1988-2008
## A summary of past studies

| *Sample Format* | *# Studies* |
|---|---|
| Single Protein | 8 |
| Single Peptide | 6 |
| Complex Mixture | 3 |
| Peptide + Protein | 2 |
| Other | 1 |

| *Study Purpose* | *# Studies* |
|---|---|
| PTM identification | 6 |
| Sequence Elucidation | 6 |
| Length of Read | 2 |
| PVDF vs Filter | 2 |
| Specific Amino Acids | 2 |
| Sensitivity | 1 |
| Sequencing Improvements | 1 |

Summarized from D.C.Brune et al."ABRF ESRG 2006 Study" Journal of BioMolecular Technologies. 18(5):306-320 (Dec 2007)

# ESRG 2008 Study Objective

To determine the effects of a homo poly amino acid N-terminal tag on Edman degradation.

Are differences observed during Edman degradation between tagged proteins a consequence of repeating amino acids in general or is there something specific about Histidines that cause variable data?

# Affinity Tags for Protein Purification

- Polypeptide sequences fused to recombinant proteins
- Allows the researcher to easily purify large amounts of recombinant protein
- Tags widely used in biotech industry
- Much variability in the type of tag, protein and purification system

# Affinity Tags: General Properties

1. One-step absorption purification
2. Minimal effect on III° Structure
3. Easy, specific removal to produce native protein
4. Simple assay of recombinant protein during purification
5. Wide applicability

# Small Peptide Affinity Tags Commonly Used

| Tag | # Residues | Sequence | Comments |
|---|---|---|---|
| Poly-Arg | 5-6 | RRRRR | - *basic*<br>- *C-term tag in bacteria*<br>- *CX purification* |
| Poly-His | 2-10 | HHHHHHHH | - *electron donor*<br>- *N- or C-term tags*<br>- *Ni-NTA purification* |
| FLAG | 8 | DYKDDDDK | - *hydrophilic*<br>- *Ab purification* |
| Strep II | 8 | WSHPQFEK | - *Biotin derivative elution* |
| c-myc | 10 | EQKLISEEDL | - *multi-system recognition*<br>- *Ab purification* |

Summarized from Terpe, K. Overview of tag protein fusions. *Appl Microbiol Biotechnology* (2003) 60:523-533

# Affinity Tags and N-terminal Sequence Analysis

Reasons

- To ensure tag was added
- Proper processing
- Difficult get N-termini from mass spec
- Correct protein, open reading frame

Challenges

- Longer Sequencing Stretches
- Repeating Amino acids
- Drop in yield of His tags
- Other proteins not having His tags bind $Ni^{2+}$ (Albumin, endogenous proteins, SLYD_ECOLI)

# Histidine Tags

- Histidines have strong interactions with immobilized metal ion matrices
- $e^-$ donor on imidazole ring form coordination bonds with transition metals (ex. $Ni^{2+}$, $Co^{2+}$)
- 6xHis tags have a high affinity to Ni-NTA and so may be easily purified
- Proteins are eluted from matrix by adjusting pH or adding free imidazole

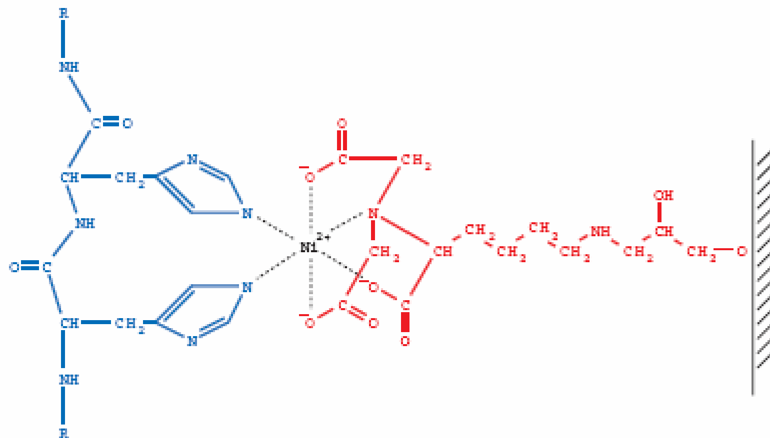


*Figure 1. Interaction between neighboring residues in the 6xHis tag and Ni-NTA matrix.*

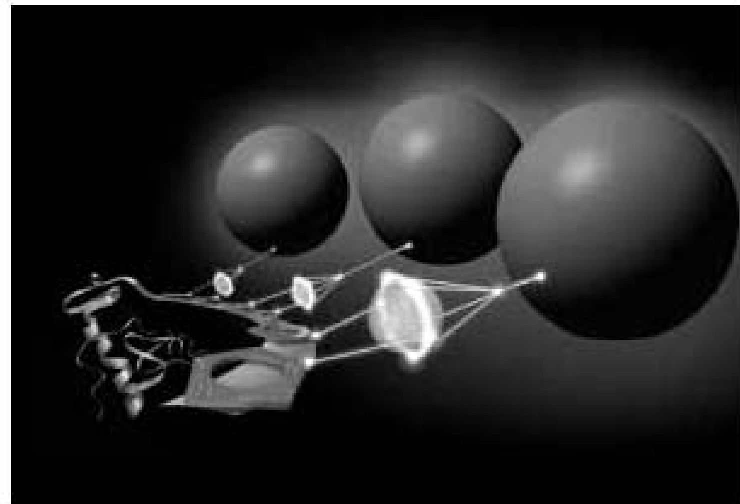


*Figure 1. Interaction between Ni-NTA and a 6xHis-tagged protein*

QIAGEN *QIAexpressionist Handbook*

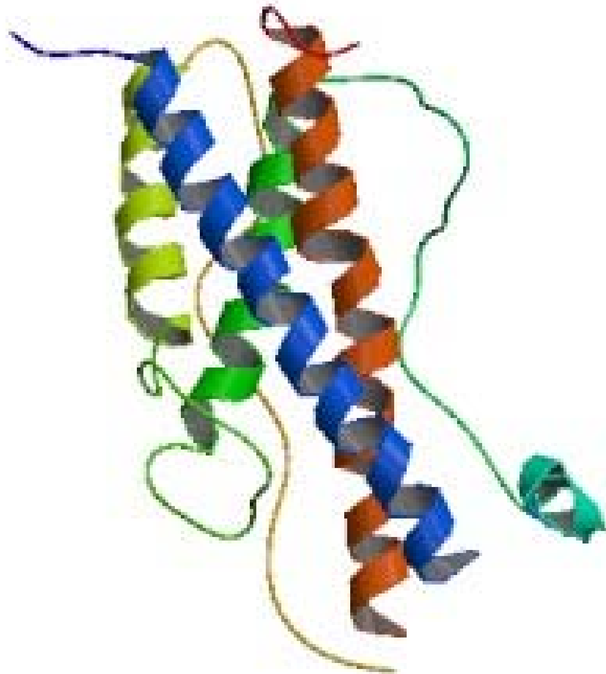
# General Cloning and Expression Protocol: Adding a His Tag

1. Insert protein DNA into a vector encoding a His-tag

2. Perform PCR with primers that have repeating His codons next to start/stop

*Modified from Wikipedia*

# Study Protein:
# Human Growth Hormone



- First recombinant pharmaceutical to be manufactured and marketed by a biotech company (1985 approval)

>P01241|SOMA_HUMAN Somatotropin - Homo sapiens (Human).
**FPTIPLSRLFDNAMLRA**HRLHQLAFDTYQEFEEAYIPKEQKY
SFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFL
RSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQ
TYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGS
CGF

- 191 Amino acids
- 2 phosphorylation sites: Ser132, Ser176
- Not an ‘ideal’ protein for Edman
- Used as real life example

Chantalat, L., Jones, N.D., Korber, F., Navaza, J., Pavlovsky, A.G. (1995) THE CRYSTAL-STRUCTURE OF WILD-TYPE GROWTH-HORMONE AT 2.5 ANGSTROM RESOLUTION. *Protein Pept.Lett.* ***2:*** *333-340*

# Sample Preparation

- Different homo-poly amino acid tags were grown on small scale (3 plates each)
- Some tags did not express even though constructs were correct!

Tags tried :

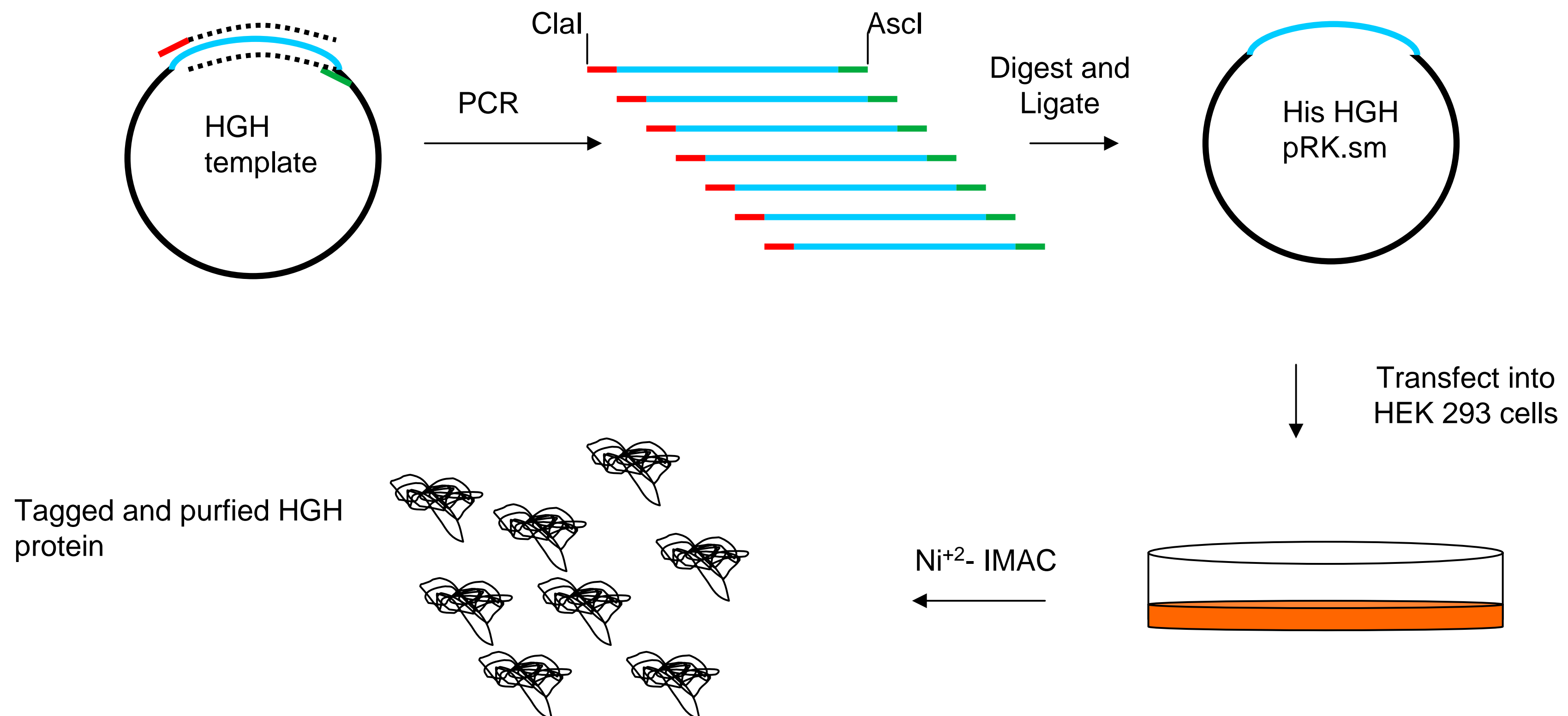
Phe - A bulky AA (like His)
- Good sequencability

Lys - charged AA (like His)

Tyr - A polar AA

Ala - Small, reliable

# ESRG samples Cloning, Expression, and Purification



# The samples…

*Sample C:*

FPTIPLSRLFDNAMLRA…

*Sample H:*

**KHHHHHHHHLE** FPTIPLSRLFDNAMLRA …

*Sample A:*

**KIDAAAAAAAA** FPTIPLSRLFDNAMLRA …

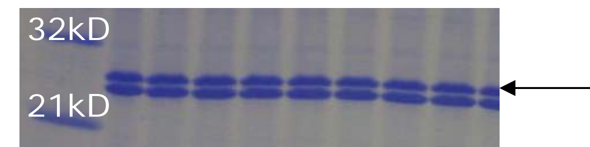
# Sample Preparation and Distribution

- Concentrations of His and Ala samples were estimated by control band intensity after coomassie staining
- ~25 pmol of each sample (Control, His tag and Ala tag) were loaded onto SDS-PAGE gels and subsequently electroblotted onto PVDF
- Two excised bands from each of 3 the samples were sent to participating labs

**Control**
**"Sample C"**

**N-term His tag**
**"Sample H"**

**N-term Ala tag**
**"Sample A"** (top band)



# Participating Lab Instructions and Samples Sent out

October 31, 2007

October 31, 2007

Dear colleagues,

Please find enclosed the ABRF-2008 ESRG samples that you requested from the ABRF Edman Sequence Research Group. This is the 20th study in an annual series designed to aid laboratories in evaluating their abilities to obtain and interpret amino acid sequence data. This year, the test samples are three expressed proteins (named H, A, and C) blotted on PVDF membrane and stained with Coomassie Blue. For each protein you will receive two PVDF pieces containing approximately 25 pmol of protein. Depending on the sensitivity of your instrument you may load either one or two of the bands for one sequencing run. The ESRG recommends loading one band on the Procise cLc or 2 bands for the Procise HT.

The object of this year's study is to ascertain the difficulty of sequencing through an N-terminal His-tag and to still be able to determine a few amino acids of the protein sequence following the tag. In order to be able to confirm the identity of the tagged proteins, the ESRG is asking all the participants to sequence 17 residues of each sample. If it is not possible for you to sequence all three samples we are asking to sequence at least samples H and A. Preferably, all raw data should be submitted in electronic form as text files created directly from the sequencing software. Make certain that you send us at least the pmol amounts of each amino acid in each cycle. On instruments connected to a Windows PC you can export the tabulated pmol raw data (open under "View", "Sequence Data") as text file (under "File", "Export"). On instruments connected to a Macintosh open the sequence data ("Analysis", "Show Sequence Data"), save the data as a text file ("File" menu select "Save As").

The ESRG is asking participants to return the raw data text files and the survey sheet by e-mail. You should have received an Excel file containing the survey sheet as attachment to the e-mailed confirmation of your sample request. If you did not receive the Excel file, or prefer a paper copy, please contact Brian Hampton (bhamp001@umaryland.edu) and he will resend it or fax you a paper copy. The files with the results should be e-mailed to Ms. Glenda Cowart (glenda.m.cowart@monsanto.com) as an attachment by December 1st, 2007. If you are returning paper copies of the survey sheet and/or raw data text files on a dos-formatted floppy disc, please mail to:

Glenda Cowart
Monsanto Co.
800 N. Lindbergh Blvd.
Mail Zone U4A
St. Louis, MO 63167, USA

In order to ensure anonymity, Glenda Cowart will remove all identifying marks prior to forwarding the data to the sequence committee for analysis. An e-mail will be sent to your facility with a three-digit code to allow you to evaluate your results as compared to others. The sequencing and survey results will be presented at ABRF 2008, February 09 - 12, 2008 in Salt Lake City, Utah, and will also help to guide future potential studies and tutorial sessions.

If your sample arrived damaged, if you have questions about the study or have problems saving the data please contact Brian Hampton at the above e-mail address. Equipment failures and "no data obtained" analyses are as important to us as data from "successful" runs. Please send us your results whatever happens. Thank you for your participation in this study!

**The deadline for receiving data for inclusion in the study is December 1, 2007.**

The Edman Sequencing Research Group:

Richard Thoma (co-chair) Monsanto.
Brian Hampton (co-chair) Univ. of Maryland
Joe Leone Pfizer

Wendy Sandoval Genentech
Klaus Linse Univ. of Texas
Peter Hunziker Univ. of Zurich
Steve Smith Univ. of Texas

Nancy Denslow (EB liaison) University of Florida

21 labs received samples

# The Results

# Results: Instrumentation and Methods Survey

ESRG 2008 Study Results: Sequence Calls, Instrumentation and Methods Survey

***addition to this survey sheet, remember to return the sequence analysis data text file with picomole valu***

| Please list any procedures used to improve the sequencability of this sample. |
|---|
| 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). |

Sequence Calls for Samples H, A and C. Include any comments or observations in the spaces provided.

| Sample H | | | Sample A | | | Sample C | | |
|---|---|---|---|---|---|---|---|---|
| Cycle | Called PTH AA | Comments | Cycle | Called PTH AA | Comments | Cycle | Called PTH AA | Comments |
| 1 | K | | 1 | K | | 1 | F | (R) |
| 2 | H | | 2 | I | | 2 | P | (R,L,V) |
| 3 | H | | 3 | D | (G,P) | 3 | T | |
| 4 | H | | 4 | A | | 4 | I | |
| 5 | H | | 5 | A | (P) | 5 | P | |
| 6 | H | (I) | 6 | A | (R) | 6 | L | |
| 7 | H | | 7 | A | (G,R,A) | 7 | S | |
| 8 | (H) | | 8 | A | (T) | 8 | R | |
| 9 | (H) | | 9 | A | (I) | 9 | L | |
| 10 | I | | 10 | A | | 10 | F | |
| 11 | E | (P,V) | 11 | A | (R,V) | 11 | D | |
| 12 | F | | 12 | F | | 12 | N | |
| 13 | P | | 13 | P | | 13 | A | |
| 14 | T | | 14 | T | | 14 | M | |
| 15 | I | | 15 | I | | 15 | L | |
| 16 | ?? | | 16 | P | | 16 | R | |
| 17 | ?? | (Y) | 17 | L | | 17 | A | |
| | | | 18 | S | | 18 | H | |
| | | | 19 | R | | 19 | R | |
| | | | 20 | (L) | | 20 | L | |

| | |
|---|---|
| Manufacturer (ex: ABI, HP) | ABI |
| Model # (e.g. 492 cLc, 494-HT, 477) | 494-HT |
| Age (in years) | 10 |
| Inst. Manufacturer Reagents ? (yes/no) | yes |
| TFA Cleavage Delivery (ex: gas, liquid) | liquid |
| Sample Support (ex: gf, PVDF) | Sample supplied on PVDF |
| Chemistry Cycle (ex: GF, PVDF peptide) | PVDF protein |
| DTT in S2B/ethyl acetate (yes/no) | No |
| Other reagents additives? Describe | No |
| Sample Solvent (ex. %TFA / %CH3CN / H2O) | Not applicable for this study |
| Number of cards loaded on sequencer | 2 |

| HPLC Information | |
|---|---|
| Column Manufacturer | ABI |
| Column Type (ex. Spheri-5 PTH, 5 um) | Spheri-5 PTH, 5um |
| Column Dimensions (ID x length, mm) | 2.1 X 220mm |
| Column Age (estimated # cycles) | 6mo. (1275) |
| Average Column Life (total cycles) | 8-10mo (> 2000) |

| HPLC Gradient | | |
|---|---|---|
| Time | % B | Flow Rate (µl/min) |
| 0 | 8 | 325 |
| 0.2 | 11 | 325 |
| 0.4 | 14 | 325 |
| 18 | 48 | 325 |
| 18.5 | 90 | 325 |
| 21.5 | 90 | 325 |
| | | |
| | | |
| | | |
| | | |

| PTH-Amino Acid Standards | | |
|---|---|---|
| Amt pmol = 4 | | |
| PTH AA | RT (min) | Peak Area |
| D | 3.93 | 193686 |
| N | 4.49 | 234778 |
| S | 5.2 | 166852 |
| Q | 5.53 | 223153 |
| T | 5.77 | 189384 |
| G | 6 | 186648 |
| E | 6.54 | 256381 |
| H | 8.09 | 227769 |
| A | 8.45 | 202255 |
| R | 10.27 | 167996 |
| Y | 10.79 | 192823 |
| P | 11.95 | 226848 |
| M | 13.68 | 217918 |
| V | 13.98 | 214019 |
| W | 16.1 | 252361 |
| F | 16.71 | 209768 |
| I | 17.11 | 219462 |
| K | 17.35 | 274763 |
| L | 17.59 | 222630 |

Information Requested:

- Sequence calls
- Instrument type (cLC or HT?)
- Reagent information
- Standard RT and areas
- Column Information
- Gradient

# Results:
## Uncorrected Raw Data Returned in Excel Format

Amino acid

Pmol value

| | ASP | ASN | SER | GLN | THR | GLY | GLU | HIS | ALA | ARG | TYR | PRO | MET | VAL | TRP | PHE | ILE | LYS | LEU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.21 | 0.31 | 0.9 | 0.55 | 0.35 | 1.31 | 0.68 | 0.73 | 0.6 | 0.57 | 0.5 | 0.43 | 0.23 | 0.43 | 0.08 | 0.38 | 0.43 | 9.44 | 0 |
| 2 | 0.63 | 0.49 | 0.93 | 0.65 | 0.3 | 1.35 | 0.61 | 5.44 | 0.39 | 0.42 | 0.53 | 0.64 | 0.09 | 0.41 | 0 | 0.56 | 0.69 | 1.03 | 0.86 |
| 3 | 0.66 | 0.43 | 1.11 | 1.06 | 0.34 | 1.22 | 0.68 | 5.5 | 0.42 | 0.48 | 0.55 | 0.74 | 0.07 | 0.56 | 0 | 0.45 | 0.38 | 0.31 | 1.27 |
| 4 | 0.69 | 0.59 | 0.96 | 0.93 | 0.5 | 1.34 | 0.93 | 5.09 | 0.54 | 0.52 | 0.73 | 0.66 | 0.08 | 0.58 | 0.05 | 0.65 | 0.63 | 0.37 | 1.31 |
| 5 | 0.91 | 0.75 | 1.1 | 0.87 | 0.42 | 1.44 | 0.89 | 5.04 | 0.5 | 0.51 | 0.68 | 0.7 | 0.26 | 0.66 | 0 | 1.24 | 0.61 | 0.34 | 1.48 |
| 6 | 0.59 | 0.73 | 1.27 | 1.1 | 0.51 | 1.48 | 1.08 | 4.91 | 0.57 | 0.56 | 0.74 | 0.75 | 0.22 | 0.7 | 0 | 1.05 | 0.61 | 0.54 | 1.72 |
| 7 | 1.13 | 0.98 | 1.53 | 1.57 | 0.64 | 1.57 | 1.4 | 4.9 | 0.61 | 0.62 | 0.71 | 0.75 | 0.26 | 0.69 | 0.07 | 1.08 | 0.93 | 0.55 | 2.2 |
| 8 | 1.18 | 1.01 | 1.52 | 1.45 | 0.92 | 1.54 | 1.41 | 4.69 | 0.73 | 0.71 | 0.78 | 0.73 | 0.3 | 0.72 | 0 | 1.41 | 1.04 | 0.59 | 2.32 |
| 9 | 1.43 | 1.11 | 1.71 | 1.64 | 0.83 | 1.67 | 1.5 | 4.96 | 0.72 | 0.78 | 1.07 | 0.81 | 0.28 | 0.85 | 0 | 1.33 | 1.04 | 0.62 | 2.58 |
| 10 | 1.24 | 0.98 | 1.67 | 1.64 | 0.77 | 1.7 | 1.72 | 2.86 | 0.72 | 0.79 | 0.99 | 0.7 | 0.35 | 0.89 | 0 | 1.37 | 1.06 | 0.7 | 6.45 |
| 11 | 1.37 | 1.01 | 1.48 | 1.53 | 0.83 | 1.65 | 4.77 | 1.5 | 0.92 | 0.78 | 0.93 | 0.79 | 0.29 | 0.91 | 0 | 1.58 | 1.08 | 0.87 | 4.85 |
| 12 | 1.4 | 1.19 | 1.83 | 1.72 | 0.97 | 1.82 | 3.56 | 0.96 | 0.95 | 0.83 | 1.07 | 0.97 | 0.34 | 0.85 | 0 | 5.24 | 1.27 | 0.87 | 3.65 |
| 13 | 1.99 | 1.31 | 2.03 | 1.77 | 1.17 | 1.95 | 2.45 | 0.64 | 0.9 | 0.43 | 1.06 | 3.01 | 0.31 | 0.92 | 0 | 3.59 | 1.29 | 0.83 | 3.19 |
| 14 | 1.88 | 1.21 | 1.81 | 1.65 | 2.51 | 1.8 | 2.05 | 0.57 | 1 | 0.9 | 1.09 | 2.74 | 0.33 | 0.97 | 0.07 | 2.33 | 1.24 | 0.8 | 3.11 |
| 15 | 1.78 | 1.22 | 1.84 | 1.75 | 2.42 | 1.96 | 1.97 | 0.51 | 1.08 | 0.9 | 1.12 | 2.13 | 0.32 | 0.96 | 0.08 | 1.86 | 3.32 | 1.05 | 3.35 |
| 16 | 1.82 | 1.2 | 1.87 | 1.8 | 1.93 | 1.96 | 1.89 | 0.52 | 1.03 | 0.57 | 1.17 | 2.58 | 0.3 | 0.99 | 0 | 1.72 | 3.32 | 1.07 | 3.42 |
| 17 | 1.87 | 1.19 | 1.85 | 1.82 | 1.46 | 1.9 | 1.89 | 0.5 | 1.03 | 1.03 | 1.33 | 2.58 | 0.23 | 1.05 | 0 | 1.63 | 2.45 | 1.01 | 4.38 |
| 18 | 1.94 | 1.46 | 2.55 | 1.98 | 1.31 | 2 | 1.86 | 0.53 | 1.11 | 1.09 | 1.24 | 2.26 | 0.36 | 1 | 0.06 | 1.58 | 1.84 | 0.93 | 4.73 |
| 19 | 1.84 | 1.34 | 2.65 | 1.96 | 1.23 | 1.94 | 1.86 | 0.49 | 1.1 | 1.32 | 1.22 | 1.79 | 0.28 | 0.92 | 0 | 1.58 | 1.52 | 0.9 | 4.59 |
| 20 | 1.92 | 1.39 | 2.53 | 2.04 | 1.19 | 1.89 | 1.75 | 0.51 | 1.08 | 1.46 | 1.24 | 1.45 | 0.32 | 0.98 | 0 | 1.6 | 1.34 | 0.88 | 4.75 |

Cycle #

# Participating Lab Information (n=21)

| | |
|---|---|
| Manufacturer and Model: | 13 ABI 494 HT (1-11 years old: average age 8 years)<br>6 ABI 492 cLC (5-11 years old: average age 7-8 years)<br>2 ABI 494 cLC (6-11 years old: average age 7 years)<br>1 ABI 491 |
| Reagents: | 18 used all instrument manufacturer reagents<br>3 used some mfg. R1, R2c, R4, R5, & Premix<br>1 said almost mfg. reagents |
| TFA Cleavage: | 19 used pulsed liquid, 3 gas phase |
| Chemistry Cycle: | 16 PVDF, 3 GFF, 1 PVDF & Prosorb, 1 pulsed Liq cLC |
| Other Additives: | 1 TCEP to R4 & R5, 1 DTT in R4 & 1 DTT in S2.<br>1 n-acetylcysteine in R5, 1 34% Me-Pip, 52%IPA in R2 |
| Bands Loaded: | 9 x 1; 12 x 2 |
| Columns Used: | 11 ABI Spherisorb 5 micron PTH column (2.1 x 220 mm)<br>1 Higgins column 3 micron (2.1 x 100)<br>1 Higgins column 5 micron (2.1 x 220)<br>8 ABI Prosice cLC PTH column 5 micron (0.8 x 250) |

Typical Gradients used:

HPLC Gradient cLC

| Time | % B | Flow Rate (µl/min) |
|---|---|---|
| 0 | 8 | 40 |
| 0.3 | 8 | 40 |
| 0.4 | 23 | 40 |
| 18 | 44 | 40 |
| 22 | 95 | 40 |
| 22.5 | 95 | 40 |
| 26 | 50 | 10 |

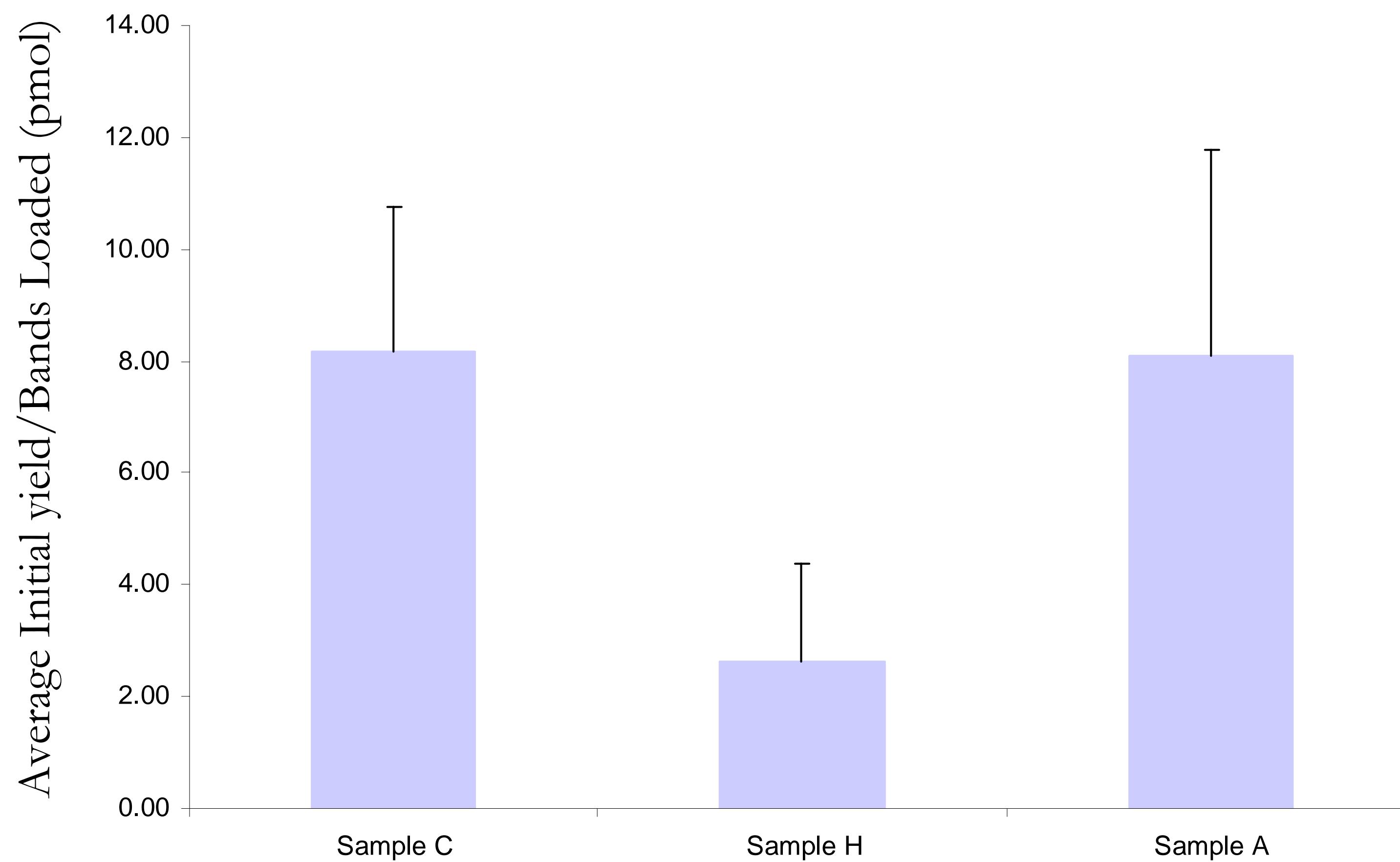
HPLC Gradient HT

| Time | % B | Flow Rate (µl/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 |

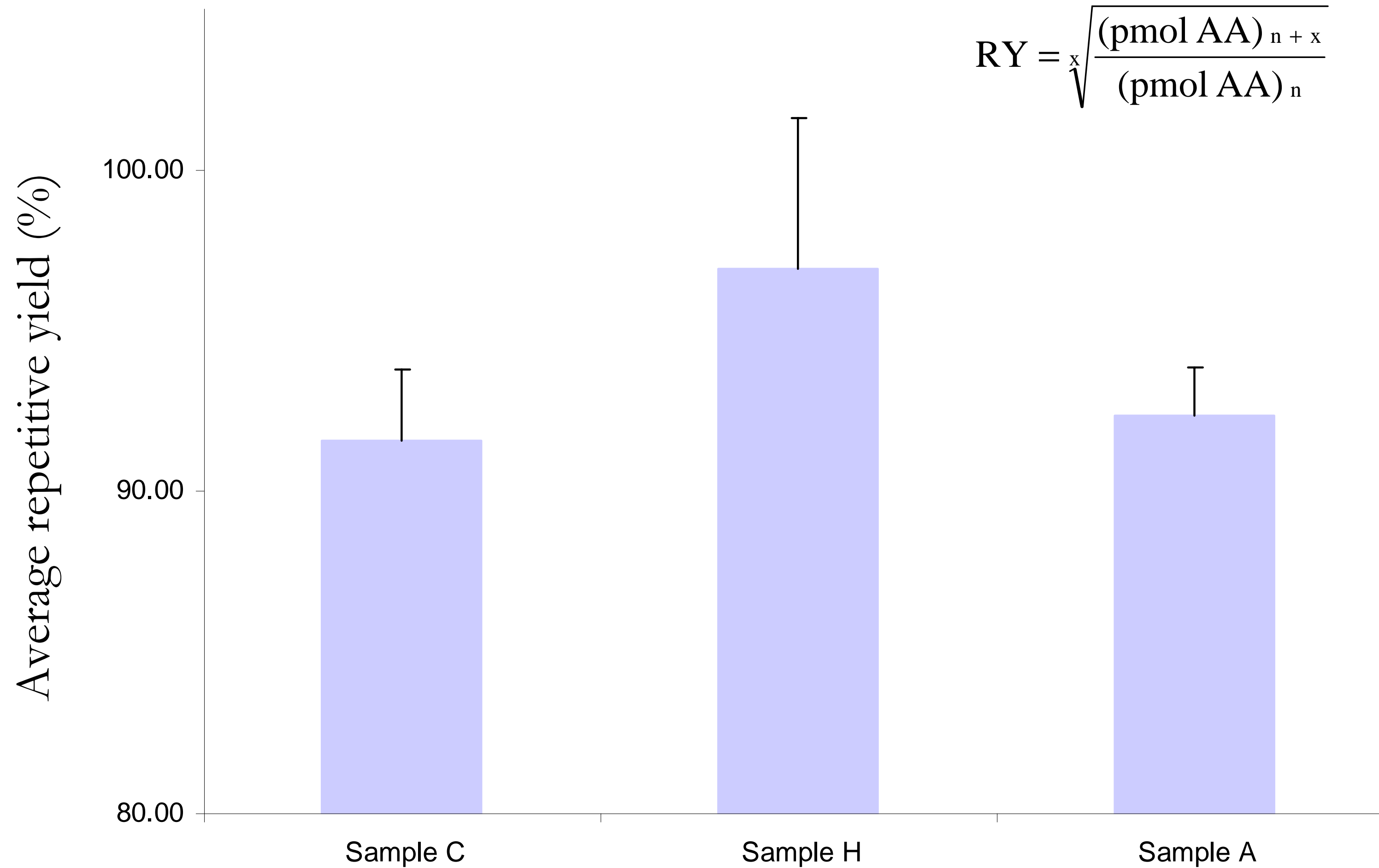
# Sequence calls and average lag (in red)

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **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 |

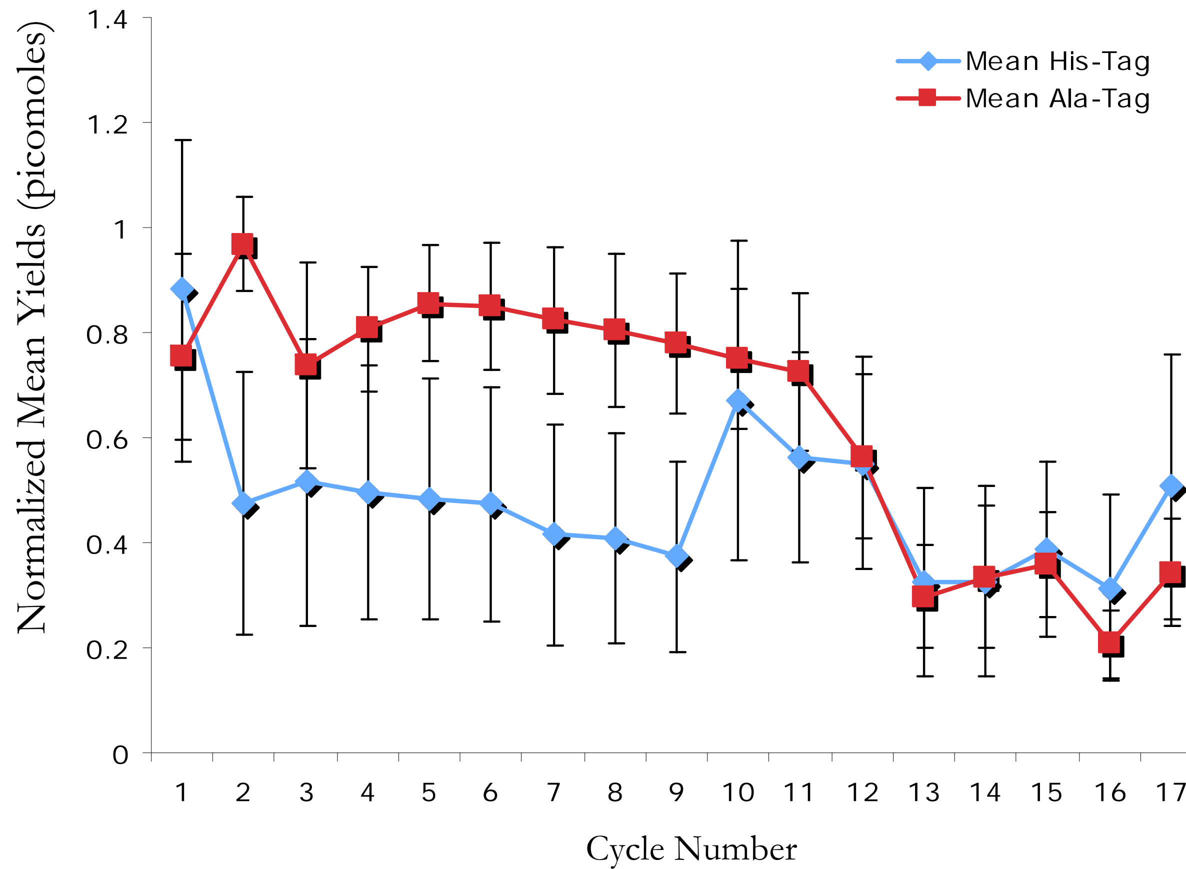
# Initial Yields



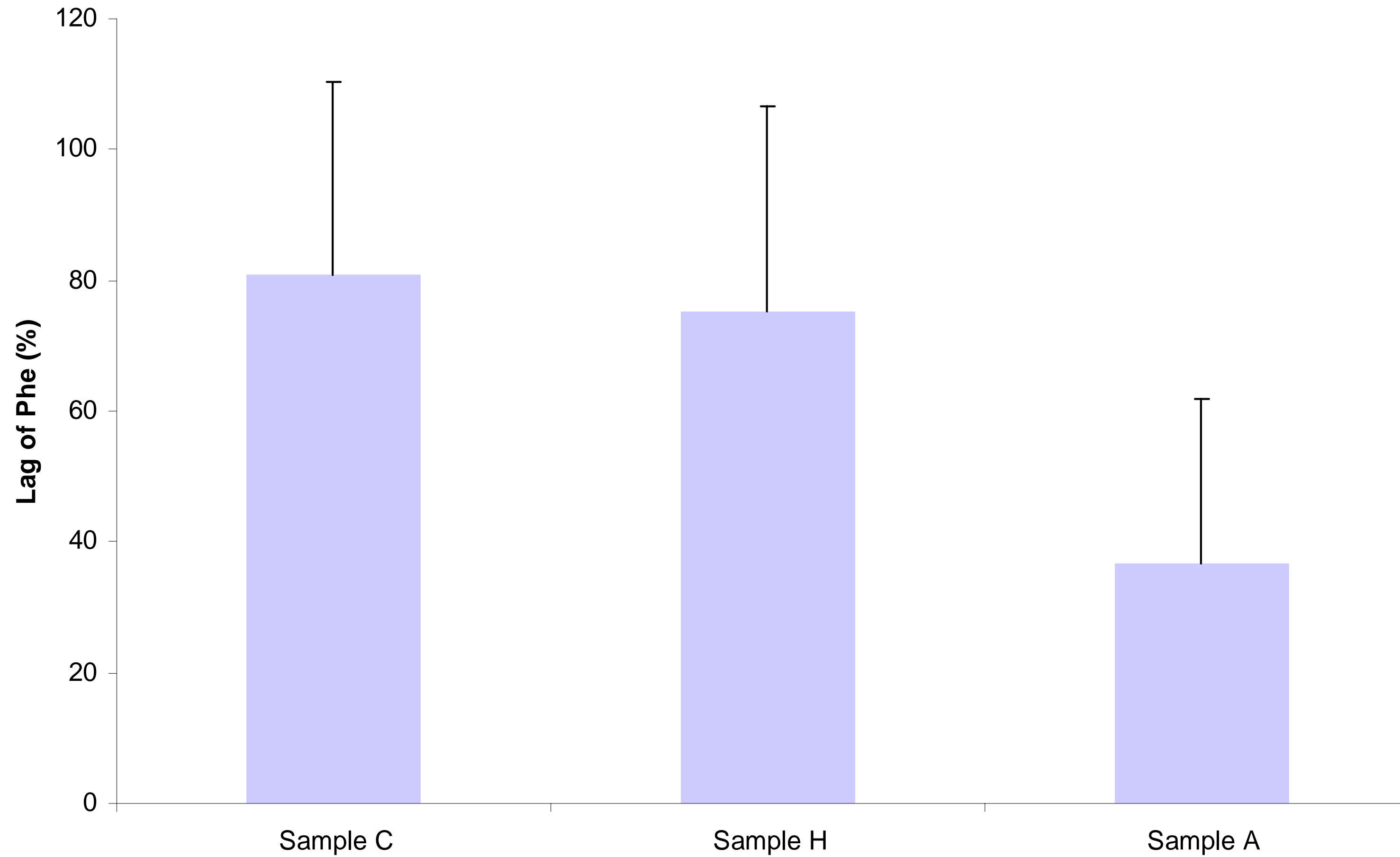
# Repetitive Yields



# Average Recoveries for His and Ala Tagged Samples



# Lag at Phe

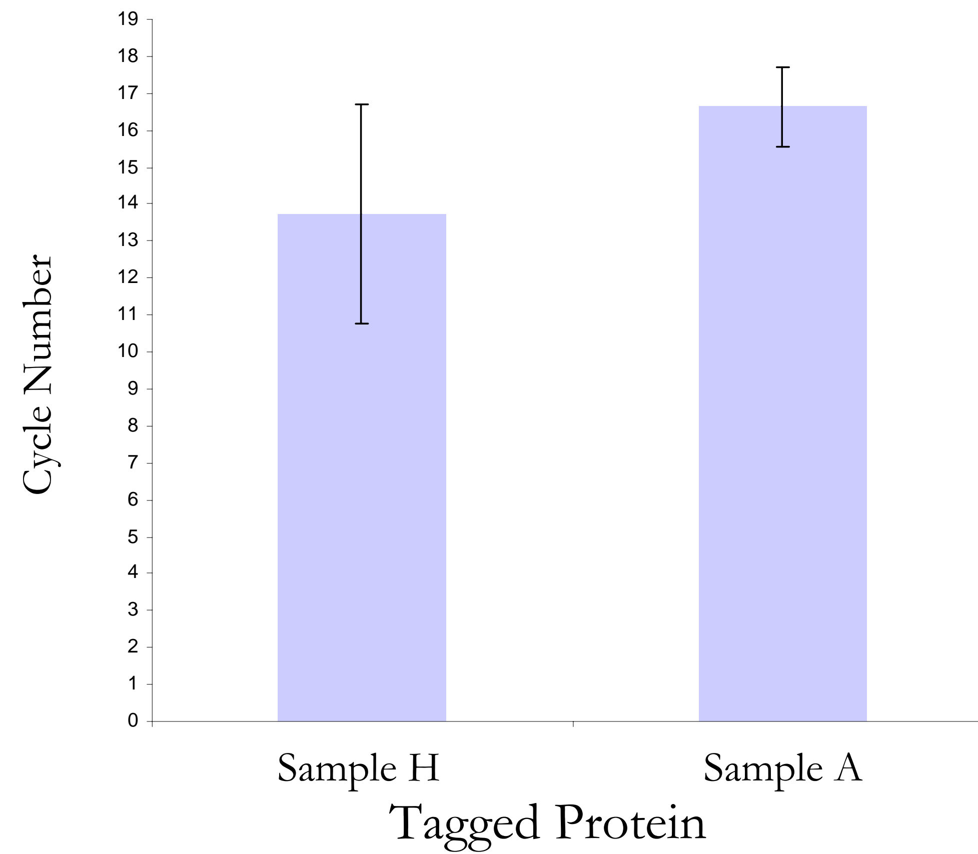



**Sample H:** K-H-H-H-H-H-H-H-H-L-E-**F**-P-T-I-P-L

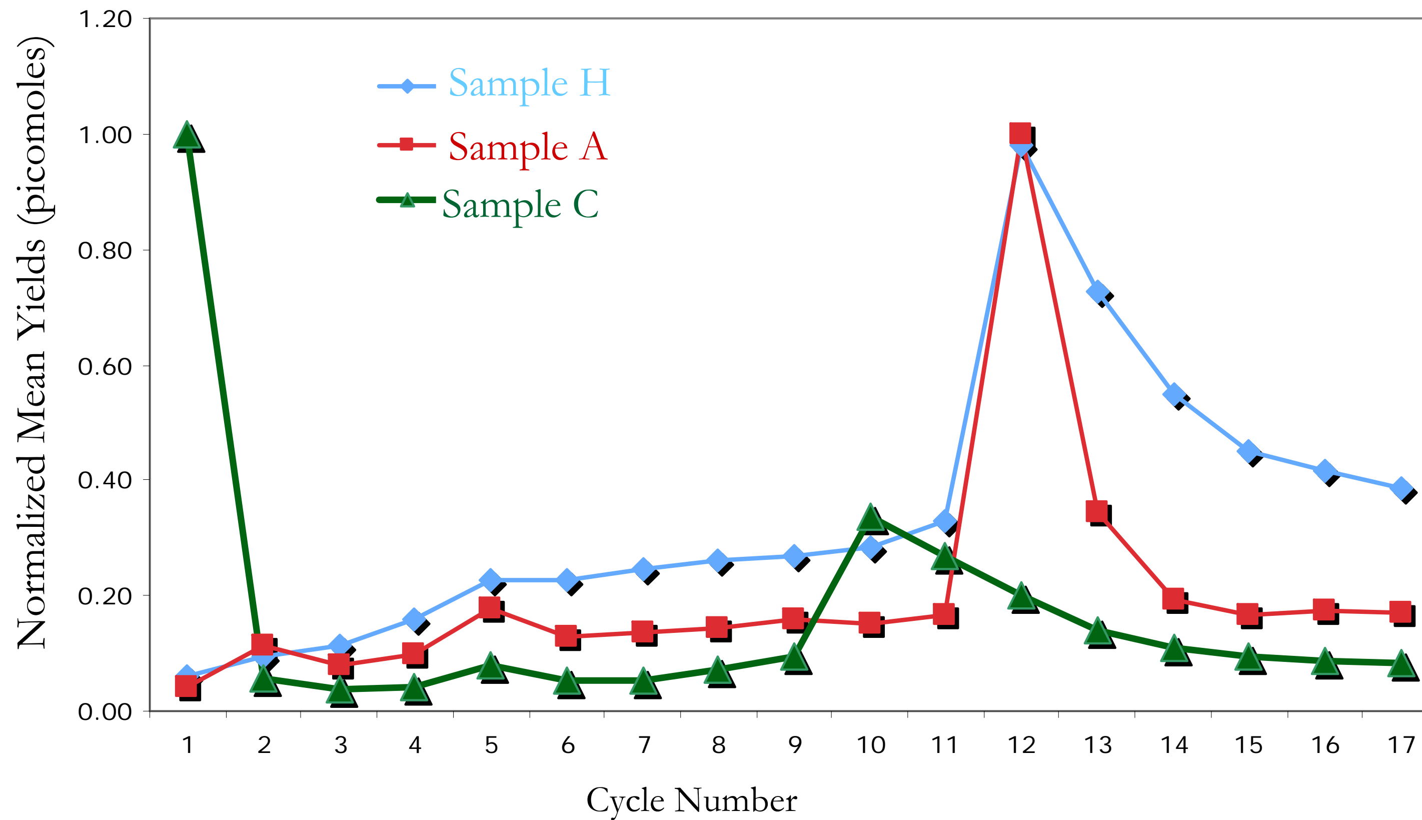
**Sample A:** K-I-D-A-A-A-A-A-A-A-A-**F**-P-T-I-P-L

**Sample C:** F-P-T-I-P-L-S-R-L-**F**-D-N-A-M-L-R-A

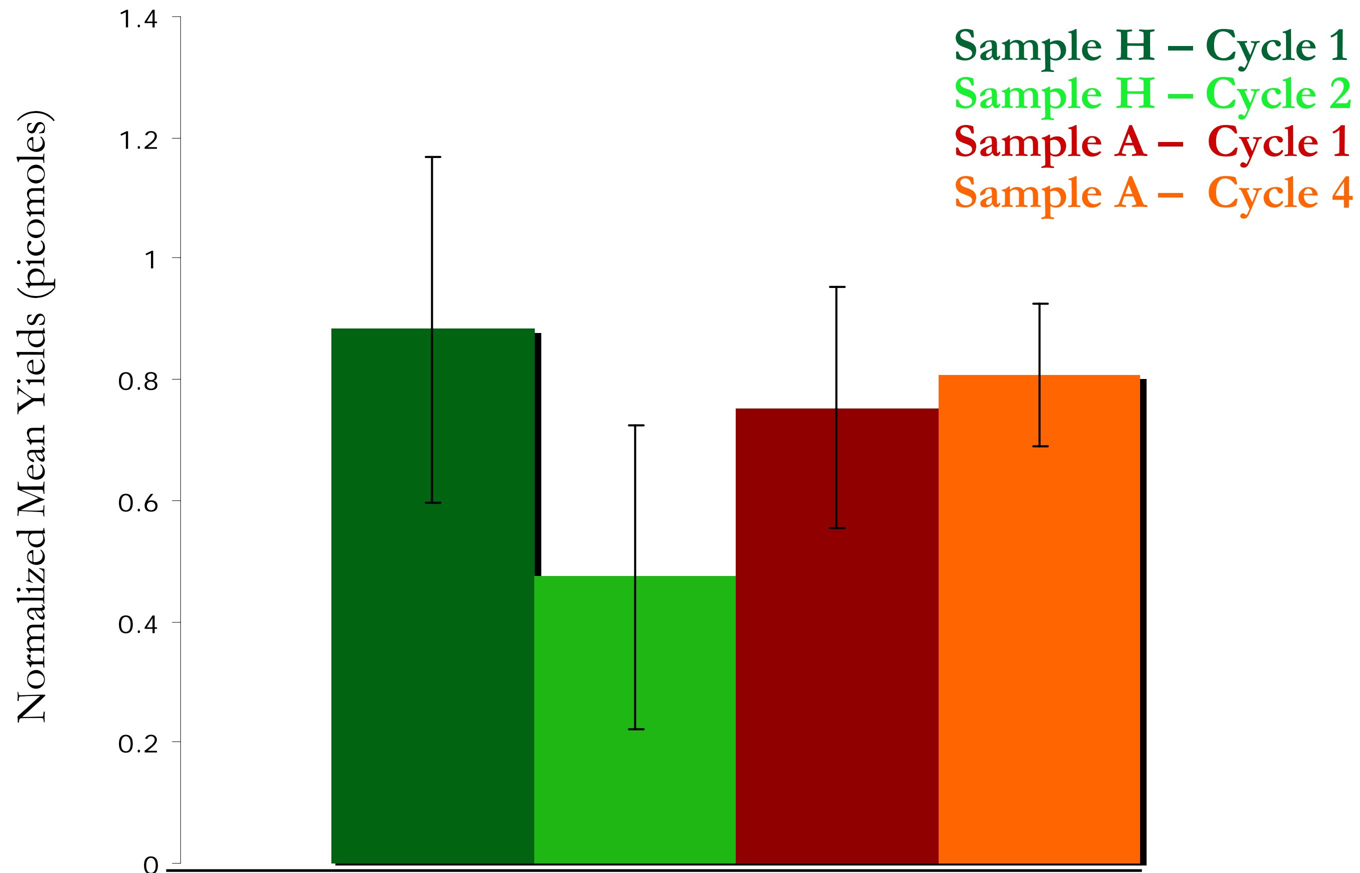
# Lag overtaking yield



# Phenylalanine Yields Throughout Run



# Decrease in pmol recovery after first cycle



# Conclusions

- Creation of a poly-amino acid is not easy to do. Other than the traditional His-tag, the only other successfully prepared poly amino acid tag for this study was an Ala tag
- There was a noticeable decrease in initial yields of the His samples versus controls
- Labs in general found it harder to call the sequence after the poly-His tag than the other two samples
- Lag was observed earlier and more consistently on the His tagged sample than the Ala tag
- High variability is due to lab to lab variability
- The majority of labs successfully sequenced seventeen cycles for all three test proteins

# Acknowledgements


- *Jessica Huard - Genentech*
  - Expression and Purification of study samples
- *Liza Ingle - Genentech*
  - Running gels, blots and sequencing for initial analysis and distribution
- *Glenda Cowart – Monsanto*
  - Accumulation of data

- *Participating labs*
- *Executive Board*
  - For support and scrutiny of study proposal
- *ESRG members*
  - Design and execution of study