

ABRF-99SEQ: Analysis of a Mixture of Peptide and Protein

Protein Sequence Research
Committee

Association of Biomolecular
Resource Facilities

ABRF Protein Sequence Research Committee Members

- Steven A. Carr
- John W. Crabb
- Gary Davis
- Karen De Jongh
- David Dupont
- Terry Lee
- Len Packman
- Linda Siconolfi-Baez
- Laurey Steinke

Karen DeJongh and Len Packman were ad hoc members during various portions of the year.

Rationale for ABRF-99SEQ

- A mixture of a peptide and a glycoprotein
- Compatible with Edman, MS/MS, and PSD sequencing
- Introduction of database searching using double sequences

Design and Synthesis of ABRF-99SEQ Peptide

- Sequence: KAWPEEESINQEFLR
- Tryptophan in third position
- Average MW = 1876.0505
Monoisotopic MW = 1874.9171
Synthesized by Fmoc chemistry
- Distributed 5 pmol

One of the goals of the committee was to present a sample which would yield results for most of the ABRF members who chose to participate in the study, but would still present some challenges for those members with advanced techniques in sequencing.

Tryptophan has historically been a difficult residue to identify.

This peptide is representative of a typical tryptic peptide.

Proline was included to introduce a lag into the sequencing, while the three consecutive glutamates would present a challenge.

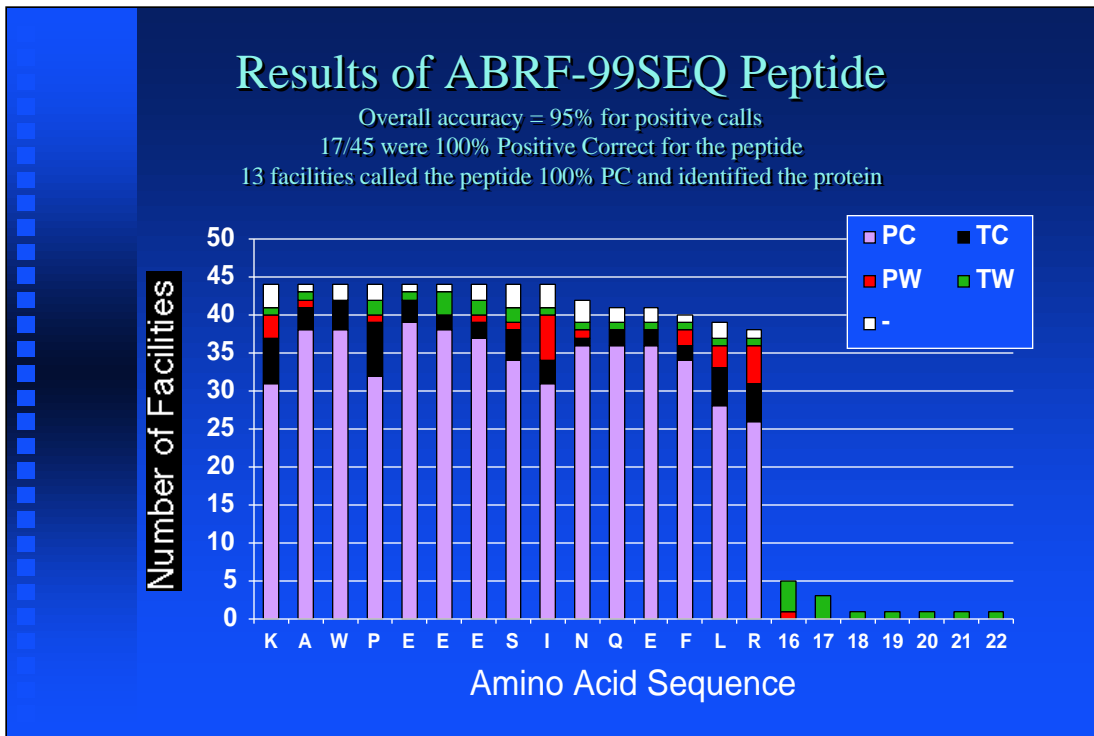
Choice of Protein: h-ATIII

- Human Antithrombin III
- A protein/peptide mix was chosen
- Protein found in the database
- N-terminal sequence:
HGSPVDICTAKPRDIPMNP MCI
- Allowed collaboration with other committees using the same protein

When the sample was sequenced by N-terminal Edman degradation, proline appears in cycle four in both the peptide and the protein. The committee thought that this might provide a challenge for those analyzing the sample.

The committee sent out procedures for interrogating databases using double sequences. We also checked, using double sequences for the peptide and the protein to make sure the protein was identified.

We placed 10 picomoles of protein into the tubes, what came out gave an average yield of less than three picomoles for the fifth cycle valine.



PC= Positive Correct

TC=Tentative Correct

PW=Positive Wrong

TW=Tentative Wrong

-- = No Assignment Made

$$\text{Positive Accuracy} = \frac{\text{PC}}{\text{PC} + \text{PW}}$$

The PSRG sorted the sequences out when amino acids were identified were correctly identified but not placed in the correct sequence . When we sorted we gave credit for a tentative correct.

Tryptophan in cycle three was not miscalled.

Isoleucine in cycle nine was the residue misidentified most frequently.

Overcalls were made by five labs. Three of the five did not determine a mass, two labs did have an accurate mass.

Identification of the Protein

- 30/45 labs were able to correctly identify the protein as Antithrombin III.
- A variety of programs were used to interrogate numerous databases.
- Positive Accuracy was very high, with only three laboratories reporting Positive Wrong calls.
- It is not correct to change a reported amino acid residue to match what is found in the database. Comparisons and corrections should be performed in comments.

Some facilities sequenced the protein past the 22 residues we gave space for in the report. Some of you can go pretty far.....

One facility, running a cLC, experienced a enough preview in the protein sequence that they called proline and valine one cycle early. Preview continued for the remainder of the cycles they ran, although not to as great an extent. Peptide was 100% correct. They still were able to ID protein, but had the wrong species.

An Appropriate Notation in the Comment Box

No residue observed, Cys by homology
with data bank search
(The sequence was marked -)

Comparison of ABRF-SEQs

	ABRF-98SEQ	ABRF-97SEQ		ABRF-99SEQ	
		Minor	Major	Peptide	Protein
Pmol distributed	2.8	2	10	5	10
Length (residues)	17	14	21	15	432
Avg. # cycles assigned	10.6	8.2	17.1	13.8	12.4
Avg. # correct (PC&TC)	8.3	4.8	14.5	12.6	11.5
Accuracy of positive calls	90.6%	71.7%	91.5%	95.4%	98.7%*
Accuracy of tentative calls	45.3%	38.0%	54.0%	62.2%	58.1%

* The reported sequence appears to have been altered to fit the database in some cases.

When we compare the results of sequencing of ABRF-97SEQ to ABRF-99SEQ it should be noted that the high peptide accuracy in 1999 is probably affected by the amount distributed. The high protein accuracy is possibly due to corrections made after the protein was identified in the database.

ABRF-99SEQ Peptide Positive Calls

- 41 of 45 labs (91%) made positive calls
- 14 of the 45 (31%) made 1-4 Positive Wrong calls
- No more than one Positive Wrong call made by any lab that correctly determined the mass

One group lost the sample in an attempt to purify it.

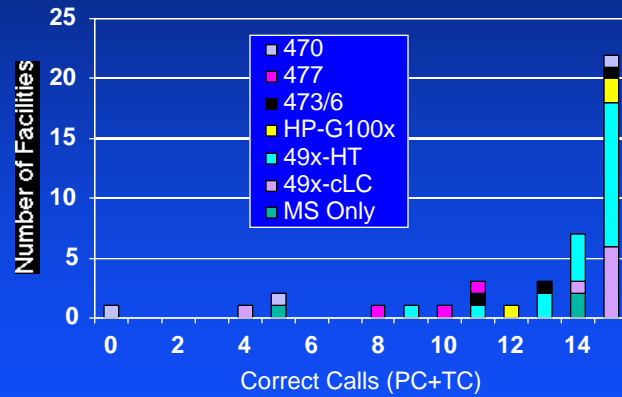
There was one group with 4 PW and one group with three PW

2.8 pmol ABRF-98 SEQ: 50% made positive wrong calls

10 pmol ABRF-97SEQB major: 45% positive wrong calls

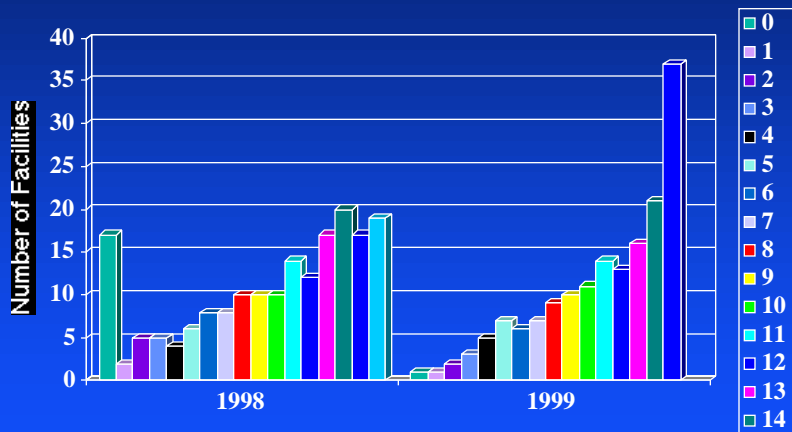
2 pmol ABRF-97SEQB minor: 40% positive wrong calls

Number of Correct Calls for the Peptide



Of the three labs which used MS/MS or PSD without Edman to determine the peptide the sequence, none were able to call the peptide sequence 100% Positive Correct.

Comparison of Number of Correct Calls 98/99



ABRF-99SEQ Best Results

- 13 labs had all 15 residues of the peptide Positive Correct and were able to identify the protein
 - ◆ 1 of 8 PE/ABI 49x-cLC
 - ◆ 10 of 20 PE/ABI 49x-HT
 - ◆ 2 of 3 HP G100x
- 4 labs identified all 15 residues PC
 - ◆ 2 of 8 PE/ABI 49x-cLC
 - ◆ 2 of 20 PE/ABI 49x-HT

All of these groups analyzed the peptide using Edman degradation.

One of the groups used PSD and one group MS/MS to assist with the analysis.

Eight of the 17 groups had a correct mass, eight had no mass, one had a mass with a large error.

One group, using an HP G1005, purified the peptide before sequencing.

Instrument Performance for ABRF-99SEQ Peptide

Edman Sequencer				
Manufacturer	Model	Avg. # Cycles Correct	Positive Accuracy	n
PE/ABI	49x-cLC	13.5	95%	8
PE/ABI	49x-HT	14.1	97%	20
Hewlett Packard	G100x	14.0	98%	3
PE/ABI	470	6.7	100%	3
PE/ABI	477	9.7	78%	3
PE/ABI	473/6	13.0	94%	4

Care should be exercised in interpreting this data because of the low number of some types of instruments represented.

The Average Number of Correct Cycles is $\text{Positive Correct} + \text{Tentative Correct}$
 $\text{Positive Accuracy} = (\text{PC}) / (\text{PC} + \text{PW})$.

Fewer 477's, HP's and 49X-HT's participated in the study than last year

Number of 470's is the same

number of cLC's is up

no data from a Porton this year.

There was one HP 241 which reported data. It is not included with the G100's

The 100% accuracy of the 4 PE 49x-cLC's last year is down. The high accuracy observed last year may have been an aberration due to the few cLC's participating.

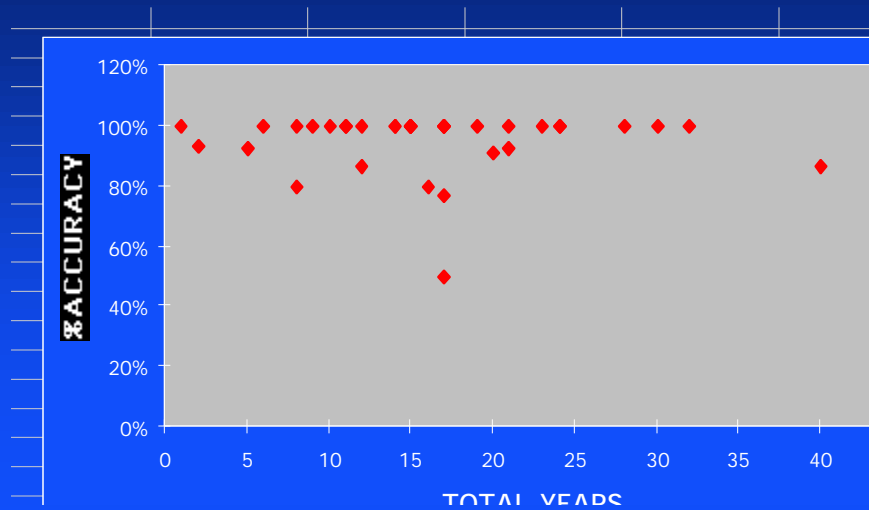
The data on the 470's is skewed. One person of the three called all 15 residues correctly (14PC, 1TC). They also had a correct mass. Two labs running 470's were unable to make any correct calls.

Instrument failures occurred on:

1 477: Reagent ran out

1 HP 241 Malfunction in two injections

% Accuracy vs. Total Years of Edman Experience



A representative of the lab with 40 years total experience came up to us at the meeting and told us that their secretary transposed two of the amino acids while typing the final report from the director's handwritten notes. Without that transposition their positive accuracy would have been 100%. It is important to double check the final report against the original data.

Survey Results for Edman Sequencing

- 41 of 45 respondents used Edman Sequencing
- 39 of 45 labs used the recommended solvent
 - ◆ Average volume was 19 μ l
 - ◆ An average of 88% was loaded
- Glass fiber filters were used by 76%, compared to 83% last year
 - ◆ 10% used bi-phasic columns (13% last year)
 - ◆ 10% used PVDF (4% last year)

The PSRG performed stability studies and recommended how to dissolve sample.

One person used a portion disc.

Of the four facilities using PVDF membrane as sample support, two had 100% Positive Accuracy for the peptide, both calling all 15 residues correctly, one had 50% Positive Accuracy for the peptide, and one person, using PVDF in a 470, was able to make no positive calls. The protein was called with 100% positive accuracy in three of the four labs. One lab reported using biobrene with the PVDF.

Determination of Peptide Mass

- This year 40% (18/45) of labs reported a value for the peptide's mass. One more lab appeared to commit sample for MS, but reported no value.
- Last year, 37.5% (21/56) of labs reported a value for the peptide's mass. Twelve more labs appeared to commit sample for MS, but reported no value.

■ We assume labs were unable to determine the mass last year due to the lower amount of peptide provide, but this is speculation. The difference could be due to the characteristics of two different peptides.

■ In summary, 40% of labs this year vs. 37.5% of labs last year determined the peptide's mass, while 42% of labs this year attempted to determine the mass vs. 59% last year.

Mass Spectrometry Results

- Eighteen of the 45 study participants (40%) used mass spectrometry.
- The average mass measurement accuracy achieved for the peptide by 13/18 labs with appropriately annotated data was 0.14 ± 0.23 Da. However, 4/18 labs did not indicate whether they were reporting monoisotopic or chemical average masses and another was in error by more than 5 Da.
- Only one lab reported a MW for the glycoprotein (56,796), a mass value within the apparent range of Antithrombin III glycoforms (56,333-57,353).

Mass Spectrometry

- Some participants using MS were not fully aware of the difference between monoisotopic and chemical average mass. The difference between an MH^+ and the MW was also not always understood.
- Three labs attempted to perform sequencing of the peptide using mass spectrometry alone. Their average Positive Accuracy was 93%.

Mass Spectrometry Cautions

- Depending on MS/MS alone to generate an entire peptide sequence is risky
 - ◆ Interpretation of the spectra needs skill and experience
 - ◆ MS/MS often does not give the entire sequence of a peptide
 - ◆ The distinction between Leu and Ile remains a difficult task

Conclusions

- Thirteen laboratories assigned all 15-amino-acid residues in ABRF-98SEQ as Positive Correct and identified the protein. Six of these used MS to determine the peptide's mass.
- There has been steady improvement in Positive Accuracy in the past three years.
 - ◆ ABRF-99SEQ (major) 5 pmol, 95%
 - ◆ ABRF-98SEQ 2.8 pmol, 91%
 - ◆ ABRF-97SEQ (major) 10 pmol, 92%
 - ◆ ABRF-97SEQ (minor) 2 pmol, 72%

■ One of the thirteen labs used a combination of Edman and MS/MS

■ This improvement in accuracy has occurred concomitantly with a replacement of 477's with newer machines. This may be a cause and effect situation.

Conclusions

- Sequence analysis of the 5 pmol peptide component of the sample mixture presented little difficulty. This may be due to improved sequencing technology.
- Thirty of the 45 participating facilities were able to identify the protein and did so with less than 3 pmol initial yield. Most groups were able to separate the peptide and protein sequences.

Conclusions

- Identifying tryptophan in cycle three was not a problem in this study.
- Nine labs performed some or all of the peptide sequence by mass spectrometry, two of which were able to call the sequence 100% Positive Correct.
- The average Positive Accuracy for facilities determining a molecular weight (97%) was slightly higher than for those not determining a molecular weight (94%).

Conclusions

- Issues facing labs performing amino acid sequencing have changed
 - ◆ Tryptophan identification
 - ◆ Amounts needed for producing reliable sequence
- The majority of the facilities taking part in this study identified a protein with less than three picomoles initial yield, and the Positive Accuracy for sequencing five picomoles of a 15 amino acid peptide was 95%.