

Title

**ABRF-00SEQ:  
SEQUENCE ANALYSIS OF A POST-TRANSLATIONALLY MODIFIED PEPTIDE**

authors

**ABRF-00 Protein Sequence Research Group**

**Steven A. Carr<sup>1</sup>, John W. Crabb<sup>2</sup>, Gary Davis<sup>3</sup>, Bryan Dunbar<sup>4</sup>,  
David Dupont<sup>5</sup>, Terry Lee<sup>6</sup>, Len Packman<sup>7</sup>, Linda Siconolfi-Baez<sup>8</sup>**

<sup>1</sup>SmithKline & Beecham, <sup>2</sup>Cleveland Clinic Foundation, <sup>3</sup>Bayer, <sup>4</sup>University of Aberdeen,  
<sup>5</sup>PE Biosystems, <sup>6</sup>Beckman Research Institute, <sup>7</sup>Cambridge University, <sup>8</sup>Albert Einstein College of Medicine

# Abstract

**This is the 13th study in an annual series designed to aid laboratories in evaluating their abilities to obtain amino acid sequence data, to determine the state of the art, and to educate ABRF members about available techniques. This year's sample, ABRF-00SEQ, was 5 pmol of a synthetic peptide that was not in the public databases. It was designed to mimic a post-translationally-modified tryptic peptide, and was chosen for compatibility with Edman sequencing, mass spectrometry, or a combination of both. Participants were asked to report anonymously the complete sequence of the peptide and to fill out a questionnaire about how the sample was analyzed. Evaluation of the data will guide the protein sequencing community as to the best methods for high-sensitivity analysis of challenging peptides.**

# Preparation of ABRF-00SEQ Peptide

Synthesis was performed on a PE Biosystems Model 433 Peptide Synthesizer using UV monitoring FastMoc™ chemistry. Fmoc-Ser(PO(OBzl)-OH)-OH was obtained from Novabiochem. Fmoc-Hyp(tBu)-OH was obtained from Bachem. Ethanedithiol and thioanisole were obtained from Aldrich. FmocArg(PMC)-resin and all other reagents, solvents and protected Fmoc-amino acids were obtained from PE Biosystems. Cleavage and final deprotection were performed in a mixture of trifluoroacetic acid (TFA):water:thioanisole:ethanedithiol (90:5:2.5:2.5). Purification was performed using a 1x10 cm Aquapore ODS column from Perkin-Elmer with a 30 minute gradient and a flow rate of 4 mL per minute. A gradient of 5% to 60% buffer B was used, where buffer A was 0.1% TFA in water and buffer B was 0.08% TFA in acetonitrile. The appropriate molecular weight of the peptide was confirmed using a PE Sciex API 100 Mass Spectrometer. Peptide concentration was determined using a PE Biosystems Model 420 Amino Acid Analyzer. Peptide sequence was verified using a PE Biosystems Procise 494.

The HPLC purified peptide solution was diluted to a final peptide concentration of 0.5 pmol/uL with 0.1% TFA / 20% acetonitrile. Aliquots (10µL) were dried in washed microcentrifuge tubes using a SpeedVac.

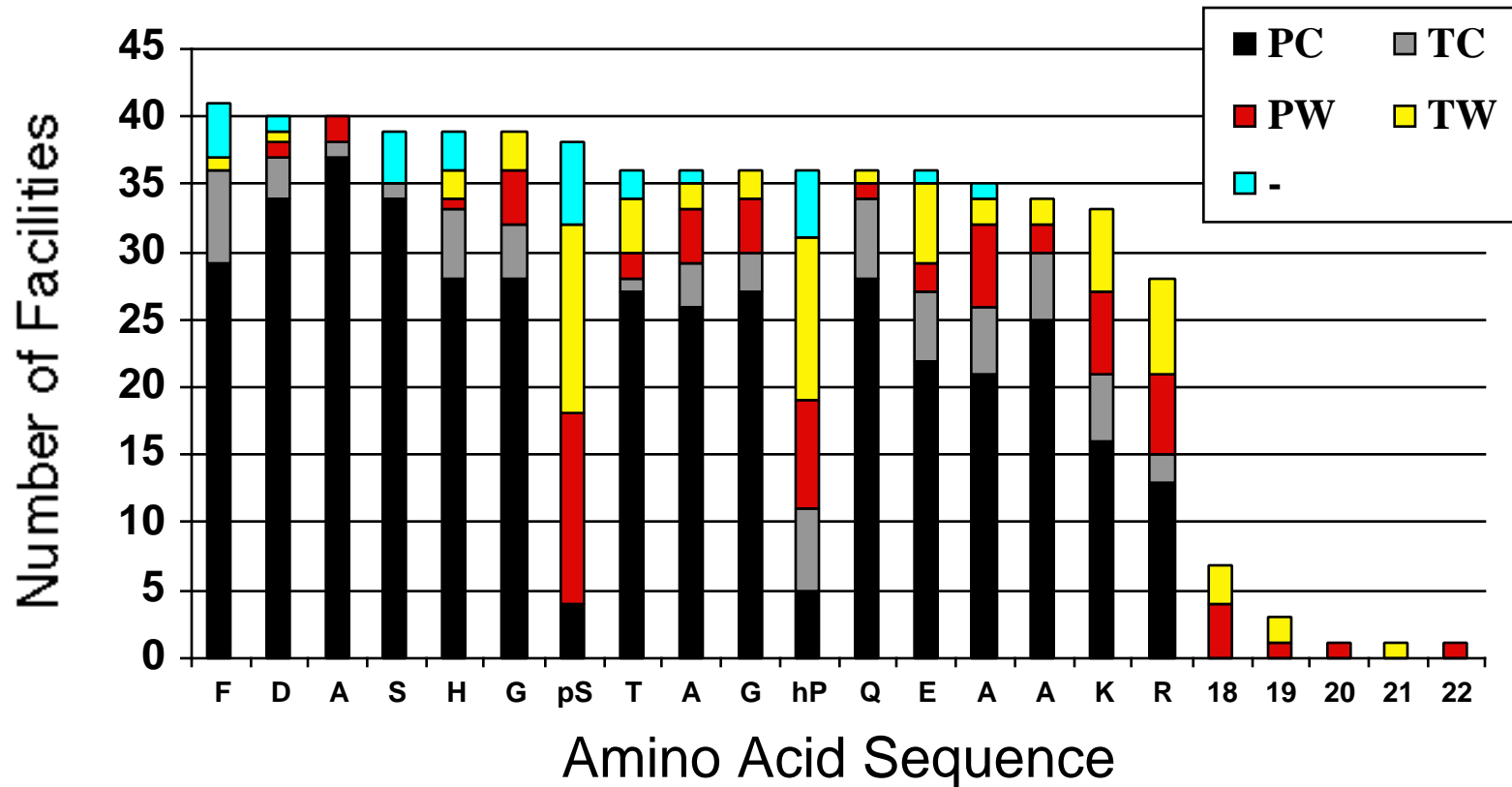
# Results of ABRF-00SEQ

Overall accuracy = 85.9% for positive calls

14/46 were 100% positive correct; none called entire sequence correctly

7 facilities reported no positive correct residues

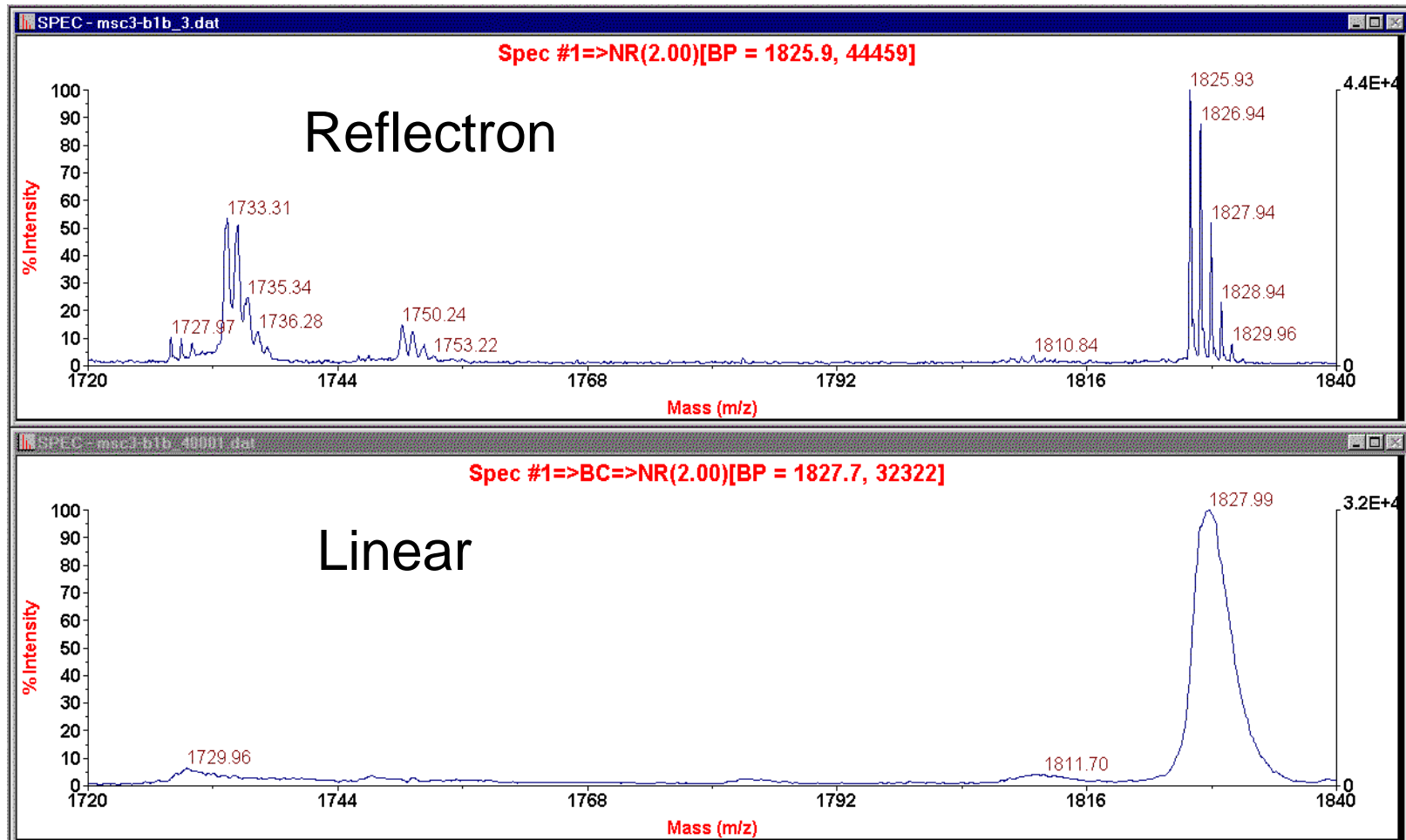
(pS = phosphoserine; hP = hydroxyproline)



# Figure 1. Results of ABRF-00SEQ

- No one called the entire sequence correctly. Four facilities positively identified 16 out of the 17 residues correctly. Two of these used only Edman and two used Edman with MS/MS.
- Seven of the 46 labs used MS/MS, and three of these also used Edman sequencing. Of the four facilities that used only MS/MS, the most positive-correct residues were 12.
- Phosphoserine at residue seven was identified only by laboratories that used MS/MS. This was the most commonly misidentified residue, and it was most often identified as Gly, which was in the previous cycle.
- Eleven out of 46 facilities identified hydroxyproline in residue 11. Three facilities that used MS/MS misidentified it as Leu/Ile, which have the same nominal mass as hydroxyproline (113).
- Six of the seven facilities that called sequence past the end of the sequence at residue 17 did not determine the mass of the peptide.
- Nine of 21 facilities that identified Lys at residue 16 did not identify any further sequence and therefore did not call Arg at residue 17.

# MALDI-TOF MS of ABRF-00SEQ



## Figure 2. MALDI-TOF MS

MALDI-TOF mass spectrometry was performed on ABRF-00SEQ, a phosphorylated peptide, using a Voyager DE-Pro in both reflectron and linear modes. The calculated monoisotopic  $MH^+$  is 1825.79 Da and the average  $MH^+$  is 1826.82 Da. The loss of masses 98 and 80 from the phosphoserine in the peptide are evident in both modes. In the linear mode, only fragmentations in the ion source are observed because fragments formed by post-source decay of metastables travel with the same velocity as the intact molecule and arrive at the detector at the same time. In the reflectron mode, both the in-source and post-source decays are observed. The post-source decay has the wrong energy to be properly focused by the reflectron, and the peaks appear at higher  $m/z$  values (1733 and 1750 vs. the expected 1728 [-98] and 1746 [-80], respectively). Note also the poor resolution of the peaks at  $m/z$  1733.3 compared to the resolution of the parent ion at  $m/z$  1825.9. A higher than normal laser power was required in order to see the -80 ions.



## Table 2

### ABRF-00SEQ MS Data Summary

Facility (Code) Amount lab usually analyses	Ioniza- tion Method & Analyzer 1	Matrix <sup>2</sup>	observed mass <sup>3</sup>	Dm from calculated mass <sup>4</sup> (Da)	Comments
2 (2013) 0.1-1 pmol	DRM ETmsms	AC	1825.78 mw1824. 8	0.007 0.021	Correct except for K being called as Q Both Edman and MS/MS data used
3 (5085) 1-10 pmol	DRM	AC	1826.69	1.097	MW only
6 (983) 0.1-10 pmol	DRM msms	AC	1825.77	0.017	Failed to note strong M-80 and M-98 ions indicating loss of phosphate. Residue 7 assigned as dehydroalanine rather than phosphoserine. MH+ was interpreted as TFA adduct to fit proposed sequence Both Edman and MS/MS data used.
8 (6731) >10 pmole	DRM <sup>5</sup>	AC	1825.84	0.053	MW only
10 (9327) 1-10 pmol	DRM <sup>5</sup>	AC	1825.83	0.054	MW only

13 (3435) 0.1-10 pmol	M(QToF) msms	AC, DHB	1825.787	0.0	Phosphate loss associated with T rather than S despite the observance of dehydroalanine in spectra. hPro-Gln incorrectly assigned as Pro-Gly-Ser. Both possibilities have the same elemental composition. The propensity for both Pro and hPro to cleave on the N-terminal side results in little or no observed fragmentation on the C-terminal side making it difficult to choose between the two possibilities. If had considered the possibility of hPro, some weak ions to support it may be in the spectra. MS only
14 (2051)	DM	AC			No mass obtained
15 (3741) 0.1-1 pmol	DRM msms ETmsms EQToF msms	AC, Nit	1825.56  mw 1824.86	0.227  0.018	HydroxyPro incorrectly assigned as Leu or Ile which have the same nominal mass as hPro. The 113 mass difference is quite prominent in the ion trap ms/ms spectra. Edman cycle was called a blank.
17 (Ag58) 1-10 pmol	DRM <sup>5</sup>	AC	1825.97	0.183	Phosphorylation noted
19 (3383) 1-10 pmol	DRM <sup>5</sup>	AC	1825.74	0.047	MW only
20 (2323) 1-10 pmol	DRM <sup>5</sup>	S	1825.79	0.003	MW only

22 (36) 1-10 pmol	DRM <sup>5</sup> EQTof msms	AC	1825.79 mw 1824.97	0.003  0.191	HydroxyPro incorrectly assigned as Leu or Ile which have the same nominal mass as hPro. The 113 mass difference is quite prominent in the QTof ms/ms spectra. Correct assignments with respect to K/Q possibilities made based on observed residue mass differences, but did not trust the data enough to call them positively. MS only
24 (6714) 1-10 pmol	EQTof msms		mw 1824.714	0.065	HydroxyPro incorrectly assigned as Leu or Ile which have the same nominal mass as hPro. The AAK part of the sequence incorrectly assigned as NR which has the same nominal mass. Based on a moderately intense ion at m/z 331.  MS only
25 (25) 1-10 pmol	DRM <sup>5</sup>	DHB	1825.744	0.043	MW only. Software problem prevented taking ms/ms data. MS only
27 (2032) 1-10 pmol	DM	AC	1825.6	0.187	MW only
28 (4151)	M	S	Avg 1829.9	3.086	MW only

31 (508) 0.1-1 pmol	DRM msms	AC	1826.28	0.493	Residue 3 assigned as dehydroalanine rather than alanine. Gly-dehydroAla incorrectly assigned as Gln (ion at 1114 not taken into account). HydroxyPro incorrectly assigned as Leu or Ile which have the same nominal mass as hPro. Y ion series not shown below m/z 400. Phosphate loss incorrectly interpreted as Val at C-Terminus. No strong B ion series to help out. Proposed sequence would have correct nominal mass.  MS and Edman data
32 (5100)	EQ				Weak ion at 609 observed, corresponds to 3+ charge state. Resolution may not have been sufficient to assign charge state.
34 (1919)	M	AC	Avg 1828.88	2.066	MW only
36 (7356)	DRM EQTof msms	AC	1825.7 1824.6	0.087 0.179	Poor Edman results complicated interpretation of the MS data. MS data system software correctly called first three residues and last eight residues.
40 (5439)	ET <sup>5</sup>		Avg 1826.0	0.194	MW only
41 (2803) >10 pmole	DRM <sup>5</sup>	AC	1825.83	0.043	MW only
44 (9573)	DRM <sup>5</sup>	AC	Avg 1824.9	1.914	MW only

46 (8057)	DRM	AC	1825.65	0.137	MW only MS attempted on a Triplequad by nanospray, but no spectrum
1-10 pmole					

1. D = delayed extraction, R = reflectron, M = Maldi on a time-of-flight analyzer, E = electrospray, T = ion trap analyzer, msms = MS/MS data obtained either by collision induced dissociation or post source decay. QTof = tandem quadrupole – time-of-flight analyzer. M(QTof) = prototype quadrupole – time-of-flight analyzer with a MALDI ion source.

2. DHB = dihydroxybenzoic acid; S = sinapinic acid; AC = alpha-cyano-4-hydroxycinnamic acid; N = Nitrocellulose

3. All values correspond to the observed monoisotopic MH<sup>+</sup> ion except for those preceded by “mw” where the value is the molecular weight calculated from one or more charge states. In one instance (#34), the average molecular weight for the MH<sup>+</sup> ion was reported.

4. For sequence FDASHG<sup>P</sup>STAG<sup>h</sup>PQAAAKR (P<sup>S</sup> = phosphoserine, hP = hydroxyproline) : monoisotopic MW = 1824.7792, (M+H)<sup>+</sup> = 1825.7870; chemical average MW = 1825.8058, (M+H)<sup>+</sup> = 1826.8137

5. Instrument had MS/MS capability but was evidently not used to obtain fragment ion data.

## Table 3: ABRF-00SEQ Survey Results Summary

### General Questions

	<u>Labs Reporting</u>	<u>Edman</u>	<u>MS</u>	<u>MS/MS</u>	<u>Edman Only</u>	<u>MS Only</u>	<u>MS and Edman</u>	<u>Edman &amp; M/MS</u>
Method	37	41	24	9	21	5	18	5
Solvent		<u>Recom.</u>	<u>Other</u>					
		35	11					
	<u>Labs Reporting</u>	<u>High</u>	<u>Low</u>	<u>Mean</u>				
Volume	33	100	5	24				
%Edman	33	100	25	78.6				
% MS MW	19	40	1	9.59				
% MS Seq	7	90	5	29.75				
% Digestion	2	40	30	35				
% Not Used	14	80	5	41.5				
No. of People	43	5	1	2.05				
No. Last Year	27							

### MS Questions

<b>MS MW</b>	<u>Labs Reporting</u>	<u>Perceptive</u>	<u>Finnigan</u>	<u>Micromass</u>	<u>HP</u>	<u>Brucker</u>	<u>Kratos</u>	<u>Prototype</u>
Instrument	24	14	1	2	1	4	1	1
Ionization	20	<u>Maldi</u>	<u>ES</u>	Some labs did both				
		21	6					
Age	24	<u>High</u>	<u>Low</u>	<u>Mean</u>				
		6	3months	2.5				
Features	21	<u>DE</u>	<u>Reflectron</u>	<u>PSD</u>	<u>Collision Cell</u>			
		18	18	14	10			
<b>MS Seq.</b>	<u>Labs Reporting</u>	<u>Perceptive</u>	<u>Finnigan</u>	<u>Sciex</u>	<u>Micromass</u>	<u>Brucker</u>	<u>Prototype</u>	
Instrument	7	1	1	2	3	1	1	
Ionization	8	<u>Maldi</u>	<u>ES</u>					
		3	5					
Age	8	<u>High</u>	<u>Low</u>	<u>Mean</u>				
		7	1	3				
Features	7	<u>DE</u>	<u>Reflectron</u>	<u>PSD</u>	<u>Collision Cell</u>			
		3	7	3	3			

Calibration	18	<u>Internal</u> 1	<u>External</u> 17				
Matrix MW	20	<u>ACH</u> 17	<u>sinnapinic</u> 1	<u>DHB</u> 2	<u>Nitrocellulose</u> 1	<u>reflectron</u> 14	
Matrix Seq	5	4		1	1	5	
MW accuracy	22	<u>&lt;0.1 Da</u> 15	<u>0.1-0.2 Da</u> 5	<u>0.2-0.5 Da</u> 3	<u>0.5-1 Da</u> 0	<u>&gt;1.0 Da</u> 3	
Routine Amt.	14	<u>&gt;10</u> 3	<u>1-10 pm</u> 11	<u>0.1-1 pm</u> 4			
No. of people	23	<u>High</u> 4	<u>Low</u> 1	<u>Mean</u> 1.83			
Experience	22	32	1	6.57	Avg per Individual		
<b>Edman Questions</b>							
Instrument	32	<u>HP</u> 2	<u>ABI</u> 37				
Model	40	<u>HP</u> 2	<u>470</u> 1	<u>473/6</u> 2	<u>477</u> 2	<u>49x</u> 25	<u>cLC</u> 7
Age	31	<u>High</u> 13	<u>Low</u> 1	<u>Mean</u> 4.11			
Support	39	<u>GFF</u> 33	<u>PVDF</u> 2	<u>biphasic</u> 2	<u>Porton Disk</u> 2		
%PTH on LC	30	<u>High</u> 100	<u>Low</u> 42	<u>Mean</u> 76.67			
Routine Amt.(a)	30	<u>&gt;10 pm</u> 4	<u>1-10 pm</u> 28	<u>0.1-1 pm</u> 5			
No. of people	39	<u>High</u> 4	<u>Low</u> 1	<u>Mean</u> 1.67			
Experience		28	3 month	8.6	Avg per Individual		

Notes

(a) One lab reported their sequencing was equally divided between two different levels

# Conclusions

- **Sequence analysis of ABRF-00SEQ (5 pmol) proved difficult. The overall accuracy for positive calls was low (85.9%). No lab called the entire sequence correctly.**
- **Seven laboratories detected a phosphorylated residue in ABRF-00SEQ and four correctly (~9%) identified pSer at residue 7 using MS/MS. For comparison, 9/74 laboratories (~12%) made a positive correct call for pSer at residue 4 in ABRF-92SEQ (37 residues, 500 pmol provided). The loss of masses 98 and 80 evident in mass spectra can facilitate the identification of phosphopeptides.**
- **About 11% of the participants (5/46) positively identified hydroxyproline at residue 11. hPro was misidentified 3x as Leu/Ile in part because these residues have the same nominal mass as hPro (113). For comparison, 34/74 facilities (46%) positively identified hPro at residue 11 of ABRF-92SEQ.**
- **This year, 52% of the participants (24 of the 46 facilities) used mass spectrometry compared to 40% last year. Only two facilities that attempted to determine the molecular weight of ABRF-00SEQ were unsuccessful. The average error in mass was similar for all instrument types (~ 0.1 Da) except for three measurements using linear MALDI-TOF MS without delayed extraction where the average error was >2 Da. The one prototype MALDI QTOF determined the mass correctly to the third decimal place.**
- **Sequence calls past the end of the peptide were minimized by facilities that determined the peptide's mass.**