

Title

**ABRF-99SEQ:
SEQUENCING A MIXTURE OF A PEPTIDE AND A GLYCOPROTEIN**

Authors

ABRF-99 Protein Sequence Research Committee

**Steven A. Carr¹, John W. Crabb², Gary Davis³, Karen De Jongh⁴,
David Dupont⁵, Terry Lee⁶, Len Packman⁷, Linda Siconolfi-Baez⁸, and Laurey Steinke⁹**

**¹SmithKline & Beecham, ²Cleveland Clinic Foundation, ³Bayer, ⁴Zymogenetics,
⁵PE Biosystems, ⁶Beckman Res Institute, ⁷Cambridge Univ, ⁸SUNY Brooklyn, ⁹Univ Nebraska Med Center**

Abstract

This the 12th study in an annual series designed to aid laboratories in evaluating their abilities to obtain amino acid sequence data. This year's sample, ABRF-99SEQ, mimicked a peptide sample "contaminated" with a glycoprotein. The peptide and protein were chosen for compatibility with sequencing by Edman degradation, mass spectrometry or a combination of both. The sample consisted of a mixture of a peptide and a protein in the five to ten picomole range. The peptide in this sample was representative of a tryptic peptide and is not found in the databases. Participants were asked to report the complete sequence of the peptide and the N-terminal sequence of the protein. Instructions were provided for using search engines with multiple entries, and the participants were asked to identify the protein. Ninety-seven samples were distributed on request to ABRF members. Analysis of the data returned by study participants will assist them in evaluating their abilities to identify proteins in the presence of a contaminant.

Purpose of ABRF-99SEQ

- To evaluate laboratory sequencing performance with a sample containing a mixture of a peptide (5 pmol) and a glycoprotein (10 pmol). Participants were asked to determine the complete sequence of a 15 residue synthetic peptide, partial N-terminal sequence of the protein and to identify the protein.
- To encourage analyses by both Edman degradation and mass spectrometry.

Materials and Methods

The ABRF-99SEQ test sample was a mixture of a synthetic peptide and a protein, human antithrombin-III. The peptide was synthesized in the lab of Gary Davis, Bayer Corporation, on a 433A peptide synthesizer (PE Biosystems) using NMP-HBTU Fmoc chemistry and pre-loaded Wang resin. The protecting groups were Lys(Boc), Trp(Boc), Glu(OtBu), Ser(tBu), Asn(trt), Gln(trt), and Arg(Pmc). Cleavage and deprotection were done in 84.6% trifluoroacetic acid, 4.4% thioanisole, 2.2% ethanedithiol, 4.4% liquefied phenol, and 4.4% water for 2 hours at room temperature. The crude peptide was precipitated, washed in t-butyl methyl ether, and purified by HPLC on a Dynamax C-18 reverse phase column using 0.1% trifluoroacetic acid with an acetonitrile gradient. The peptide was dried and reconstituted in water. Its concentration was determined by Karen West, from the Cleveland Clinic Foundation, using PTC amino acid analysis and automatic HCl vapor phase hydrolysis (Model 420H analyzer, PE Biosystems).

Antithrombin-III was a gift of Dr. Tim Edmunds, Genzyme Corporation, Framingham, MA. After dialysis into PBS, its concentration was determined by Dean McNulty at SmithKline Beecham using ninhydrin amino acid analysis on a Beckman 6300.

Sample components were mixed together and final concentrations adjusted with water to 0.5 pmol/ μ l peptide and 1.0 pmol/ μ l Antithrombin-III. Aliquots (10 μ l) were lyophilized in Eppendorf polypropylene tubes previously washed with water and methanol. Prior to distribution, committee members analyzed the sample using Edman degradation and mass spectrometry. A total of 97 samples were mailed to those who requested them.

Results of ABRF-99SEQ Peptide

Peptide Overall accuracy ~95% for positive calls

17/45 were 100% positive correct

13 facilities called the peptide 100% PC and identified the protein

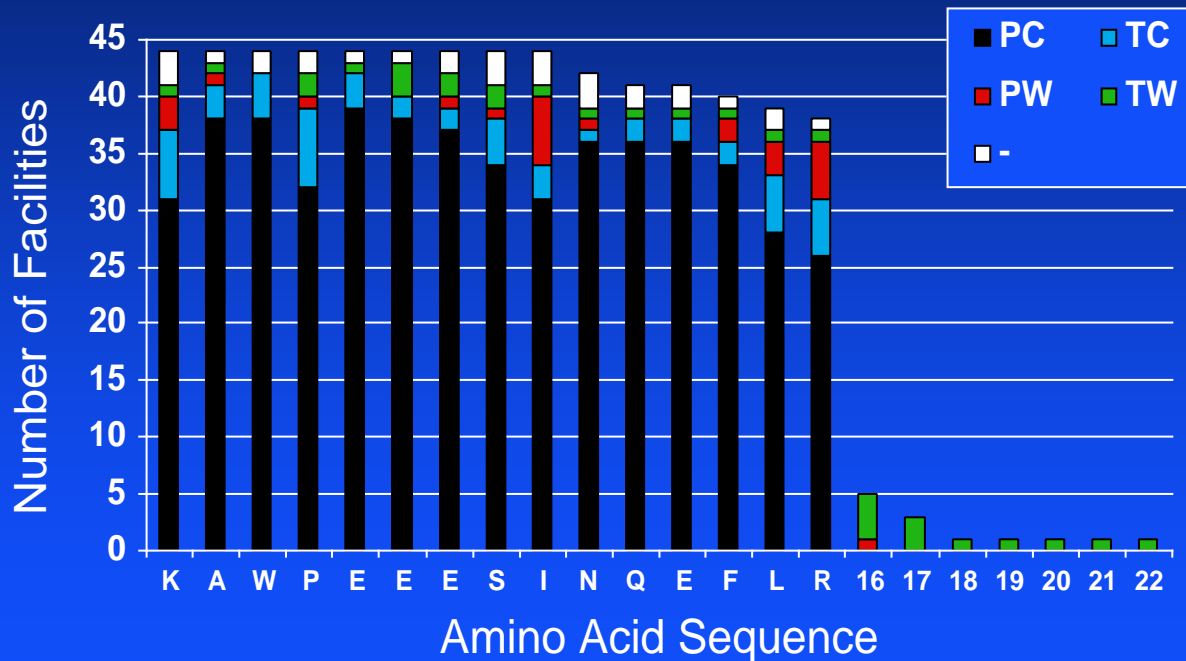


Figure 1. Results of ABRF-99SEQ Peptide

Tryptophan used to be considered a difficult amino acid to identify by Edman degradation. In this year's study there were no wrong calls for tryptophan in residue three and 42 out of 45 labs identified it correctly.

Isoleucine in residue nine was the residue misidentified most frequently. Arginine in residue 15, the C-terminus of the peptide, was the residue with the fewest number of Positive Correct calls; it was also misidentified frequently.

The average amount of alanine in cycle two of the peptide (corrected for the amount loaded onto the sequencer) was 3.2 picomoles. This is an average initial yield of 64%.

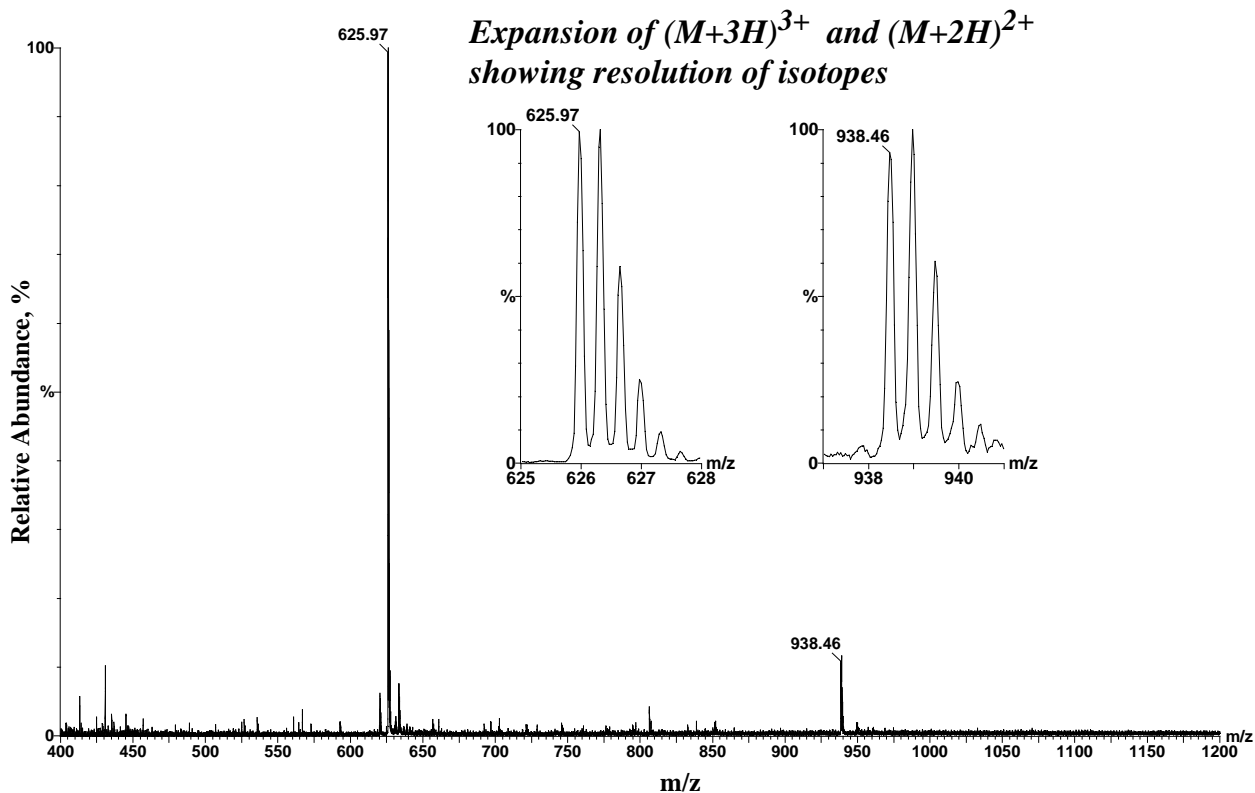
Figure 2. Comparison of ES-MS/MS and PSD Derived Sequence Data

ES-MS/MS (**panel A**) of the $(M+2H)^{2+}$ of the ABRF-99SEQ peptide (a few hundred fmol) yielded a complete series from y_2 to y_{14} and b_2 to b_{14} (**panels B and C**). The data shown were obtained using a commercial hybrid quadrupole-time-of-flight mass spectrometer (Micromass Q-TOF) and an ion trap (Finnigan LCQ). B-ions (including multiply-charged b-ions) are more abundant in the ion trap data than in the quadrupole data. The Q-TOF data are very similar to the data obtained from a triple quadrupole with respect to the ion series observed and the relative abundances of the ions (data not shown). The principal differences are the much higher resolution and mass accuracy obtained with the quadrupole-TOF analyzer. PSD-MALDI of the $(M+H)^+$ gave an analytically useful spectrum containing many of the same ions, but there are some significant differences (**panel D**). The signal-to-noise ratio of the PSD data is lower than that obtained by ES-MS/MS which makes distinguishing real ion signals from noise more difficult. Interpretation of the low mass region of the PSD data was complicated by the presence of abundant "internal" ions, the strongest of which began with the Pro at position 4 and extending toward the C-terminus (e.g., PE, PEE, PEEE, PEEES). These ions are observed to only a minor extent in the ES-MS/MS data for this peptide, although they can sometimes be quite intense. The b_4/y_{11} pair are very weak in the PSD data. It is common to get low or no cleavage C-terminal to Pro when fragmenting singly-charged precursors.

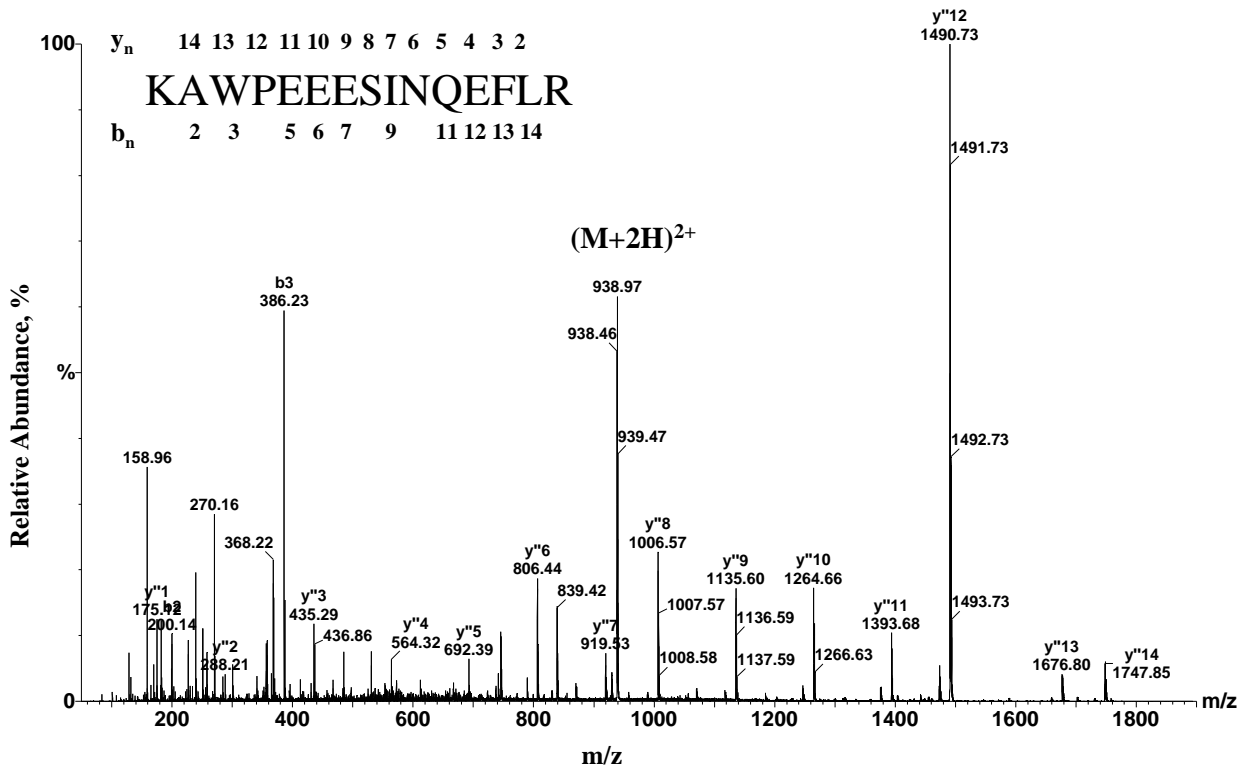
Some problems with the use of MS for peptide sequencing remain. Leu and Ile cannot be distinguished under the low-energy MS/MS conditions commonly used. In addition, Gln and Lys require either accurate mass measurement or chemical derivatization (e.g., acetylation) to be distinguished by MS. Conventional quadrupole mass analyzers and ion traps do not have sufficient mass accuracy to make this distinction. However, the mass accuracy afforded by hybrid quadrupole-TOF instruments is sufficient to distinguish Gln from Lys which differ in their accurate masses by 0.036 Da.

The Committee expresses their gratitude to Dr. Susan Chen, Dr. Xiaolong Zhang and Mr. Michael Huddleston, Research MS Laboratory, Department of Physical and Structural Chemistry, SmithKline Beecham Pharmaceuticals for help with acquiring and interpreting the MS data shown above.

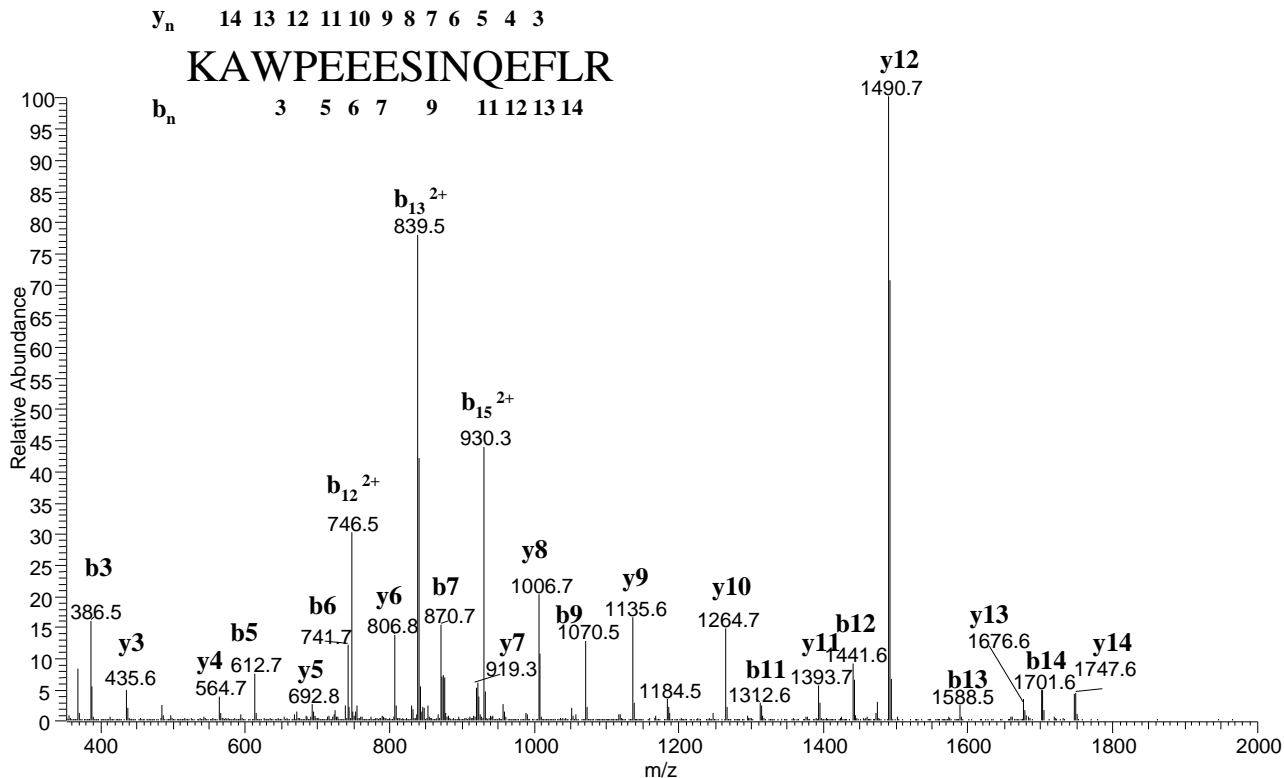
A. Electrospray MS of ABRF99 Peptide Using Quadrupole-TOF



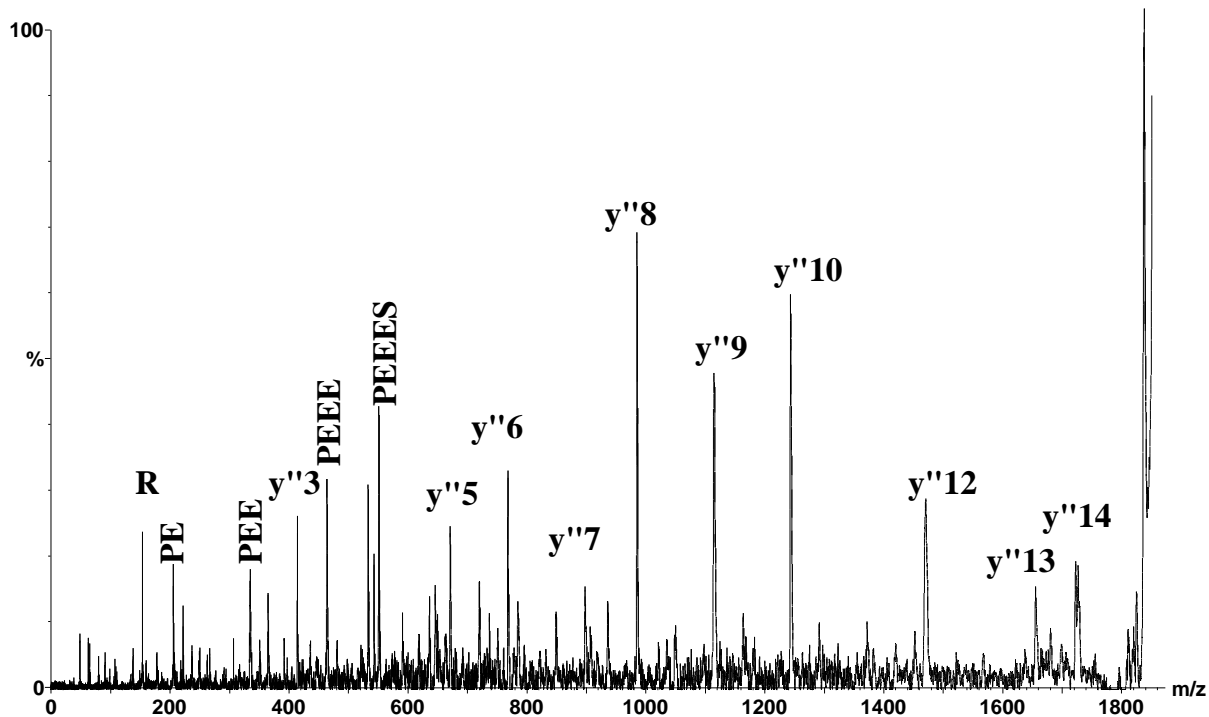
B. ES-MS/MS of $(M+2H)^{2+}$ Using Quadrupole-TOF



C. ES-MS/MS of $(M+2H)^{2+}$ Using Ion Trap



D. PSD-MALDI-TOF MS of ABRF99 Peptide (M+H)⁺



PC	TC	PW	TW	-	Edman Instr	pmol Val%	% for Edman	MS Instr	Edman N-term	Enzyme	Comment	ID	Program	Database	Accuracy (Positive)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
468	50	6	36	85												18	25	22	19	30	30	30	1	29	30	29	26	25	26	28	22	26	21	11	11	9		
5	3	3	6	2												5	3	3	6	2	2	3	1	1	1	2	3	1	3	2	5	2	1	1	2	1		
1	1	1	2													1	1	1	2																			
2	2	3	3	2												2	3	3	2	1	1	6	2	1	1	2	3	4	2	2	2	2	1	3	1	2		
8	5	5	4													8	5	5	4				1	26	1	1	2	3	4	2	2	3	1	3	3	8		
Total:	468	50	6	36	85				41			30		Protein Sequence:	H	G	S	P	V	D	I	C	T	A	K	P	R	D	I	P	M	N	P	M	C	I		
Fac #	PC	TC	PW	TW	-	Edman Instr	pmol Val%	% for Edman	MS Instr	Edman N-term	Enzyme	Comment	ID	Program	Database	Accuracy (Positive)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	15	4			1	476	0.6	75%	V DE STR	+		+	BLAST	NRDB	100%	h	G	s	p	v	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
2	16	1		2	1	476	0.3	50%	V DE RP	+		+	MS-EDMAN	SWISS PROT	100%	-	G	S	p	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
3	14			2	1	49X	0.4	100%		+		+	BLAST	EMBL	100%	H	G	S	d	V	D	I	g	T	A	K	P	R	D	I	-	M	N	P	M			
4	16			6	2	49X	1.0	100%		+		+	PEP SCAN	SWISS PROT	100%	-	-	-	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
5	20			2	3	49X	2.3	100%		+		+	FASTF	NRDB	100%	H	G	S	-	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
6	15	2		2	3	49X	0.7	80%	MM ES	+		+				100%	H	G	S	-	V	D	I	e	T	A	K	P	R	D	I	P	M	N	P	M		q
7	17			1	1	HP G1005	1.0	80%		+		+	FASTA	SWISS PROT	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
8	18	1		1	1	476	1.6	100%		+		+	MS-EDMAN	SWISS PROT	100%	H	G	S	p	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
9	14			3	1	49X	0.6	80%		+		+	MS-EDMAN	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	-	M	N	P	M			
10	20			2	2	49X	1.1	94%	HP G 2025	+		+	BLAST	SWISS PROT	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
11					3	49X	nd	37%	V DE RP	+			Not Detected																									
12	15			3	3	477	1.7	100%		+		+	MS-EDMAN	SWISS PROT	100%	-	G	S	p	V	D	I	-	T	A	-	P	R	D	I	P	M	N	P	M			
13	13	4		3	3	49X	nd	95%	B BIFLEX	+		+	FASTA	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
14	12			2	3	49X	0.5	100%		+		+	MS-EDMAN	SWISS PROT	86%	-	-	P	V	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
15	15			3	2	HP 241	0.7	95%	V DE RP	+		+	BLAST	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	-	P	M						
16		10		2	1	470	1.0	100%		+		+	MS-EDMAN	OWL		d	g	s	p	v	d	i	q	t	a	k	p											
17						nd			B BIFLEX			Sample Lost																										
18	18	1		1	1	470	3.0	90%	nd	+		+	FAST A	SWISS PROT	100%	h	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
19	18				1	49X	nd	100%		+		+	BLAST	SWISS PROT	100%	H	G	S	P	V	D	I	C	T	A	K	P	R	D	I	P	M	N	P	M			
20	10	3		1	4	49X	1.0	100%		+		+	MS-EDMAN	SWISS PROT	100%	-	g	-	-	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
21	20			2	2	49X	1.3	95%	MM ZE	+		+	SCAN PROSITE	SWISS PROT	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
22	8			10	1	49X	1.2	50%	F LCO	+	TRYP SIN	ID by pep mass ES MS/MS	+	SEQUEST	OWL	100%	-	-	-	-	V	D	I	-	T	A	K	-	-	D	I	-	-	-	-	-		
23						49X	nd	98%	V DE STR	+			Not Detected																									
24						49X	nd	95%	MM ZE	+			Not Detected																									
25						470	nd	100%		+			Not Detected																									
26	18			1	3	49X	2.5			+		+	FAST A	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
27						nd			V DE RP		TRYP SIN	ID by pep mass MALDI MS	+	MS-FIT	NRDB																							
28	20			2	1	HP G1005	1.2	33%		+		+	BLAST	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
29		5		4	1	49X	2.8	100%		+		+				h	g	p	p	v	i	i	t	a														
30	9	8		2	2	49X	0.8	50%	V DE STR	+		+	MOTIF	SWISS PROT	100%	h	G	s	P	V	D	I	-	T	A	K	p	-	d	i	p	m	n	p				
31	18			3	1	49X	6.4	100%		+		+	BLAST	ISREC	100%	-	G	d	P	V	D	I	s	T	A	K	P	R	D	I	P	M	N	P	M		r	
32	20			2	2	49X	1.4	95%	V DE RP	+		+	BLAST	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
33	17			1	1	49X	3.0	100%		+		+	PROWL	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
34	17			1	1	49X	0.4	90%		+		+	FAST A	SWISS PROT	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
35						nd			S M-QTOF		TRYP SIN	ID by pep mass MALDI MS	+	PEP SEARCH	SWISS PROT																							
36						49X	nd	100%		+			Not Detected																									
37	5	2		2	1	6	49X	0.5	100%	+						71%	-	-	-	N	V	D	I	-	P	A	K	-	-	d	i	end						
38		5		2	10	5	49X	nd	100%	+				PROT INFO	NRDB		*	*	k	g	g	-	g	-	g	m	k	p	d	a	i	p	d	-	g	-	-	
39						HP G1005	nd	100%		+			Not Detected																									
40	15			1	1	49X	1.1	70%		+		+	BLAST	SWISS PROT	100%	H	G	S	v	V	D	I	-	T	A	K	P	R	D	I	P	M						
41		3		7	10	477	nd			+			Not Detected			s	-	-	-	I	D	I	n	-	-	I	-	-	-	-	g	y	-	v				
42						477	nd	100%		+			Not Detected																									
43						nd			V DE STR				Not Attempted																									
44	18			1	1	473	nd	100%		+		+	BLAST	NRDB	100%	H	G	S	P	V	D	I	-	T	A	K	P	R	D	I	P	M	N	P	M			
45	17	1				49X	0.7	90%	V DE RP	+		+	FIND PTRN	NRDB	100%	H	G	S	P	V	D	I	c	T	A	K	P	R	D	I	P	M	N	P	M			
Legend												MS Instr Abbreviations				* The committee offset the sequence from facility 38 by two residues. Their sequence began kgg...																						
Fac#: Facility number assigned by the committee												B BIFLEX				Lowercase=tentative call																						
PC: Positive Correct												F LCO				BLUE: Ambiguity of multiple residues resolved by the committee																						
TC: Tentative Correct												HP G 2025				RED: Positive Wrong																						
PW: Positive Wrong												K M 4																										
TW: Tentative Wrong												MM ES																										
-: No call												MM 2E																										
Edman Instr: Edman sequencer model												S M-QTOF																										
pmol ValS = Yield of PTH-Val at cycle 5 (nd = no data provided)												V DE STR																										
% for Edman: % of sample used for Edman sequence												V DE RP																										
MS Instr: Mass spectrometer model (nd = no data provided)																																						
Edman N-term: Edman sequencing performed																																						
Enzyme: Enzyme used for digestion																																						
Comment: Comment																																						
ID: Correct identification of protein																																						
Program: Database searched program used to identify protein																																						
Database: Databases searched																																						
Accuracy (Positive) = Positive Correct/(Positive Correct + Positive Wrong) = PC / (PC + PW)																																						

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

(Legend to Table 3-Instrument performance)

Facilities using ABI Procise 49X-HT and HP G1005 instruments were able to call the peptide sequence with better than 97% Positive Accuracy ($\frac{PC}{[PC+PW]}$) and an average of 14 amino acids called correctly ($\frac{[PC+TC]}{\text{number of instruments}}$).

The Positive Accuracy of 100% on the ABI 470s is skewed. One site of the three did an excellent job, calling all 15 residues correctly (14 PC and 1 TC). This site reported a correct mass. Two were unable to make any correct calls.

None of the four facilities using only MS/MS or PSD were able to call the peptide sequence 100% positive correctly.

There are fewer ABI 477, Hewlett Packard and ABI 49X-HT sequencers in the study this year as compared to last year. The number of 470s remained the same as last year while the number of ABI Procise 49x-cLCs is greater than last year. This year no participants used a Beckman or Porton sequencer.

Last year, for the 2.8 pmol sample, the ABI Procise 49x-cLC had a markedly better average number of correct cycles (16.8 out of 17) and Positive Accuracy (100%) than any other sequencer used in the study. This year, for the 5 pmol peptide sample, the Positive Accuracy from the ABI 49x-cLC was slightly lower than that of the Hewlett Packard G1005 or ABI Procise 49x-HT instruments. Average number of correct calls was also slightly lower.

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.

Table 5 Mass Spectrometry Data Summary for ABRF-99SEQ

Facility(amount lab usually analyses)	Ionization Method & Analyzer ¹	Matrix ²	observed mass ³ (Da)	m from calculated mass ⁴	Comments
1 (1-10 pmol for MS)	DE-RE-M	AC	1874.96m	.04 Da	MALDI-PSD used to sequence peptide (also did Edman); had correct interpretation including assignment of internals and immonium ions; minor quibble: y-ion labeling in the data table started from wrong end of the peptide.
2 (1-10 pmol for MS)	DE-RE-M	AC (pep) S (prot)	1875.75m	0.83 Da high ?	There is either a labeling or a calibration error in the MALDI MW data for the peptide, the value is too far off given use of reflectron which should have yielded monoisotopic mass value. An error in either the data or the analysis of the Edman run resulted in the insertion of an additional Ala at position 15. They did not sufficiently trust MS data. The extra Ala is inconsistent with both the observed MW and their assignment of the y ion series in the PSD spectrum. Only group to successfully obtain MW of protein by MS (MALDI in SA: obs = 56796).
6	ES				no mass value reported
10	M	S	1870.8a	5.3 Da; 0.3%	
11	DE-M	AC	1876.04a ?	0.01 Da ?	

13	DE-RE-M	AC	1874.82m	0.1 Da	
15	DE-M ⁵	AC	1875.7a ?	?	
17	DE-RE-M	AC	1875.27m	0.35 Da	MS MW only; lost sample during attempt to purify
18	M ⁶	?	1876.13a ?	0.08 Da ?	MS MW data used to resolve ambiguity in first Edman cycle and to assign Arg as the C-terminal residue.
21	DE-RE-M	AC	1874.90m	0.02 Da	
22 (1-10 pmol req. for MS)	DE-M ⁵ ES-IT (nanospray)	S	1876.6a 1874.8m	0.55 Da 0.1 Da	PSD obtained; masses obtained consistent with Edman-derived sequence; claimed observation of y11, but data not shown Sample cleaned up via 1 x 0.33 mm C8 column using 1% formic acid/MeOH to elute into nanospray needle. ES-MS/MS of peptide and protein digest: y2-y14; b3 – b14 observed. All masses reported for peptide are within 0.3 Da of calculated monoisotopic value. SEQUEST analysis of ES-MS/MS data for the protein digest gave 23 peptide IDs, 12 matching hATIII.
23 (1-10 pmol req. for MS)	DE-RE-M	AC	1874.7m	0.2 Da	Obtained PSD spectrum of good quality. Most ions IDed correctly, except for strong internal series (PE, PEE, PEEE, PEEES, PEEESI). Cleavage C-terminal to Pro is highly unfavorable (no y11 obs).
24	DE-RE-M	AC	1874.936m	0.02 Da	Contamination in Edman caused three calls for N-terminal AA, MS data consistent with only K or Q. MS data used to call Arg at end of sequence.
27	DE-RE-M	AC	1874.88m	0.04 Da	MS data only. Attempted PSD, but low data quality contributed to misinterpretation – called first 3 residues correctly; rest wrong.

30	DE-RE-M	AC	1874.90m	0.02 Da	used MW data to sort out ambiguity in Edman cycle 1 and to assign Arg as final residue. Acquired ES-MS/MS data but did not discuss or appear to use.
32	M ⁵ (noDE)	AC	1875.5a ?	0.6 Da ? (0.03%)	Program used to calculate MW for peptide gave incorrect value.
35	RE-M and RE-M- MS/MS on Q- TOF ⁷	DHB	1874.90m	0.02 Da	MS data only. MS/MS of 1+ and 2+ ions produced by MALDI showed interesting differences: 2+ gave yn ions almost exclusively (including y11; few low mass bn), whereas 1+ gave mix of bn and yn (no y11, same as for PSD) with yn dominating; Did not indicate how they differentiated between K and Q – got K1 wrong. Properly calibrated, this instrument should be able to distinguish between K and Q based on the ca. 36 mmu difference between the two. Protein IDed by tryptic mass mapping and DB search. Excellent data quality (resolution, mass accuracy and sensitivity).
43	DE-RE-M ES-IT nanospray	AC	1874.9m	0.0 Da	MS data only. Sequenced peptide by ES MS/MS; sequence correct except for ambiguities of L/I and K/Q, although they correctly assigned K1 based on low abundance b-17 ions; Ambiguity at L9 was not noted in results sheet.
45	DE-RE-M	AC	1875.00m	0.08 Da	

Table 5 Legend

1. ES = electrospray on quadrupole; ES-IT = ES on ion trap; ES-TOF = ES on an orthogonal time-of-flight; M = linear MALDI-TOF; DE-M = delayed extraction (or similar) linear MALDI-TOF; RE-TOF = reflectron MALDI-TOF; DE-RE-TOF = delayed extraction with RE-M; M-MS/MS = MALDI MS/MS with collision cell.
2. DHB = dihydroxybenzoic acid; S = sinapinic acid; AC = alpha-cyano-4-hydroxycinnamic acid
3. Some laboratories reported MH⁺ observed, while others reported molecular weight (MW) by subtracting 1 from the observed mass. Values shown are determined MW; in cases where the laboratory did not specify whether the reported value was MW or MH⁺, and we had to try to determine this from other information provided (e.g., linear MALDI yields average mass), the masses are **shown in bold blue**; if we could not use other information, and the assignment is ambiguous, the value is followed by a question mark; m= monoisotopic; a = chemical average; ? = could not be determined if the masses reported corresponded to either MH or MW, and/or to average or monoisotopic values.
4. A question mark following the value indicates that the difference shown is based on our assumption of whether the lab reported MH or MW; a question mark alone indicates we could not determine a difference (see footnote 3). For sequence KAWPEEESINQEFLR: monoisotopic MW = 1874.9161; (M+H)_m⁺ = 1875.9240; chemical average MW = 1876.0505; (M+H)_{ave}⁺ = 1877.0584;
5. Had reflectron, but did not use it for this analysis.
6. No data on instrument or methods given.
7. Prototype instrument, not commercially available.

ABRF Seq99 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>
Method	45	41	19	7	26	4	15
Solvent		<u>Recom.</u>	<u>Other</u>				
		39	4				
	<u>Labs Reporting</u>	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>	
Volume	43	100	5	19.37	20	20	
%Edman	40	100	37	88.21	98.75	100	
% MS MW	17	20	2	8.09	5	5	
% MS Seq	8	30	6	13.25	10	10	
% Digestion	3	70	10	36.67	30	NA	
% Not Used	9	98	7	45.33	30	30	
No. of People	43	4	1	1.81	2	1	

MS Questions

	<u>Labs Reporting</u>	<u>Perceptive</u>	<u>Kratos</u>	<u>Micromass</u>	<u>HP</u>	<u>Brucker</u>	<u>Prototype</u>	
MS MW								
Instrument	18	10	1	3	1	2	1	
Ionization	18	<u>Maldi</u>	<u>ES</u>					
		17	1					
Age	18	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>		
		5	1 week	2.03	2	1		
Features	17	<u>DE</u>	<u>Reflectron</u>	<u>PSD</u>				
		14	16	14				
MS Seq.		<u>Perceptive</u>	<u>Finnigan</u>	<u>Prototype</u>				
Instrument	7	3	3	1				
Ionization	7	<u>Maldi</u>	<u>ES</u>					
		4	3					
Age	7	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>		
		3	0.5	1.36	1	1		
Features	4	<u>DE</u>	<u>Reflectron</u>	<u>PSD</u>				
		3	4	3				
General								
Calibration	17	<u>Internal</u>	<u>External</u>					
		4	13					
Matrix MW	17	<u>ACH</u>	<u>sinnapinic</u>	<u>DHB</u>	<u>reflectron</u>			
Matrix Seq	5	14	2	1	11			
MW accuracy	16	<u><0.1 Da</u>	<u>0.1-0.5 Da</u>	<u>0.5-1 Da</u>	<u>>1.0 Da</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>
		9	2	3	2	0.57	0.06	0.02
Routine Amt.	8	<u>1-10 pm</u>	<u>0.1-1 pm</u>					
		6	2					
No. of people	14	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>		
Experience		3	1	1.86	2	1		
		25	0.5	6.7	5	5		

Edman Questions

Instrument	40	<u>HP</u>	<u>ABI</u>				
		4	36				
Model	40	4	<u>470</u>	<u>473/6</u>	<u>477</u>	<u>49x</u>	<u>cLC</u>
			3	3	2	20	8
Age	39	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>	
		15	0.2	4.02	3	3	
Support	39	<u>GFF</u>	<u>PVDF</u>	<u>biphasic</u>	<u>Porton Disk</u>		
		30	4	4	1		
Mod. Cycles		7					
%PTH on LC	36	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>	
		90	50	71.19	70	80	
Routine Amt.	42	<u>10-100 pm</u>	<u>1-10 pm</u>	<u>0.1-1 pm</u>			
		9	27	6			
No. of people	39	<u>High</u>	<u>Low</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>	
Experience		3	1	1.69	2	2	
		31	1 month	9.62	9	12	

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.
- Identifying tryptophan in cycle 3 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.
- 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).
- The average Positive Accuracy for facilities determining a molecular weight (97%) was slightly higher than for those not determining a molecular weight (94%).
- Some participants using MS were not fully aware of the difference between monoisotopic and chemical average mass, or when it is possible and appropriate to use one or the other, and how this decision is affected by the resolution achieved by the mass spectrometer. The difference between an MH+ and the MW was not always understood. [The meaning and usage of these terms is discussed in Mass Spectrometry in the Health and Life Sciences (Ed. A.L. Burlingame and S.A. Carr) Humana Press, Totowa, NJ 1996, Appendix XI; and in Current Protocols in Protein Science, (J.E. Coligan, et al. Eds.), Chapter 16.1.1].