ABRF-PRG05: *De Novo* Peptide Sequence Determination

C.W. Turck[1], A.M. Falick[2], J.A. Kowalak[3], W.S. Lane[4], T.A. Neubert[5], B.S. Phinney[6], S.T. Weintraub[7], and K.A. West[8]

[1]Max Planck Institute of Psychiatry, Munich, Germany; [2]HHMI, University of California, Berkeley CA; [3]National Institutes of Health (NIMH), Bethesda MD; [4]Harvard University, Cambridge MA; [5]New York University School of Medicine, New York, NY; [6]Michigan State University, East Lansing MI; [7]The University of Texas Health Science Center at San Antonio, TX; [8]Galson Laboratories, East Syracuse NY

GOAL

The 2005 Proteomics Research Group (PRG) study was designed to assist ABRF members to evaluate their proficiency in sequencing unknown peptides that are not included in any published database. From a comparison of the results obtained by different strategies, participating laboratories will be able to gauge their own capabilities and establish realistic expectations for the approaches that were used.

ABSTRACT

A common request of proteomics core facilities is protein identification. However, in some instances primary sequence information for the protein in question is not present in public databases. In other cases, the amino acid sequence of a protein may differ in some way from the predicted sequence in the database as a result of gene mutation, gene splicing, and/or multiple posttranslational modifications. Thus, it may be necessary to determine the sequence of one or more peptides *de novo* to identify and/or adequately characterize the protein of interest. The primary goal of this study was to give participating laboratories an opportunity to evaluate their peptide sequencing capabilities. Samples containing 3-6 pmol each of five synthetic peptides with amino acid sequences not present in public databases were sent to 106 participating laboratories. At least one non-standard amino acid was present. A detailed analysis of the sequencing results is presented below.

INTRODUCTION

Proteomics core laboratories are often presented with unknown proteins to be identified. Sometimes, these are not identifiable by commonly-used strategies that involve tryptic digestion and database searching. There are several reasons why a peptide might not be identified by means of a standard database search using mass spectrometric data. It might be modified in some way that is not being considered by the database search program being used, it might not have a required sequence characteristic (*e.g.*, no *C*-terminal K or R from a tryptic digest), or it might come from an organism for which the primary sequence is not known. Sometimes a homologous protein can be identified, but this requires that the sequences have a high degree of identity. For example, if an unknown protein is 95% identical to a known one, there is a greater than 60% probability that a 20-residue peptide from the unknown protein will have at least one substitution compared to the corresponding known peptide. A direct sequencing approach may then be the only viable approach.

The primary goal of the 2005 PRG study was to give each participating laboratory a chance to evaluate its capabilities and practices in the following areas:

Peptide sequencing

Methods for the identification of unusual amino acids

Use of software to assist in the interpretation of de novo sequence data

The sequences of the peptides synthesized for this study are shown in Table 1. No specific approaches were recommended, although it was anticipated that tandem mass spectrometry and possibly Edman sequencing would be employed. Each of the laboratories that requested a sample was provided with a mixture consisting of 3-6 pmol each of five synthetic peptides which had sequences that were not present in any public database. The sample was supplied as a dried pellet that could be dissolved in most common aqueous solutions; one peptide (A1) proved somewhat difficult to dissolve. As with any "real-life" sample, there were minor contaminants present. There was either a K or an R at the *C*-terminus of each peptide, analogous to tryptic peptides; one peptide had a double "missed cleavage" and another contained hydroxyproline. Participants were asked to return experimental evidence for each sequence they determined, along with a completed web-based questionnaire.

METHODS

Synthesis The peptides were synthesized and purified at the following locations: A1, A2 and A3 at the HHMI Mass Spectrometry Laboratory at UC Berkeley; T50 at the NYU Protein Chemistry Laboratory; and J1 at the Macromolecular Structure Facility, Michigan State University. The synthetic peptides were analyzed by reversed-phase HPLC and MALDI-TOF mass spectrometry to verify purity.

Composition analysis - Amino acid analysis was conducted on small portions of A2, A3 and T50, individually dissolved in the appropriate volume of water to yield 1 mg/mL stock solutions. For each of these three peptides, 3 µL of the stock solution was added to an amino acid analysis tube. The blank contained 3 µL of 1% acetic acid. The samples were dried in a vacuum centrifuge, sealed and analyzed in duplicate for amino acid content using a Waters AccQtag precolumn AAA in conjunction with a Waters 2690 HPLC equipped with a Waters 2475 fluorometer.

For distribution to requesting laboratories, 3-6 pmol of each peptide was added to a 0.5-mL polypropylene tube and the peptide mixtures were dried in a vacuum centrifuge. Dried samples were sent to 76 laboratories in North America, 20 in Europe and 10 in other countries.

Table 1. Amino acid sequences of the five peptides present in the unknown sample

Peptide	No.	M_r (Da)	MH^+	Sequence
T50	1	1192.8276	1193.8349	LGAILKKLIPK
A2	2	1395.6610	1396.6683	AYTFNMGQHSLK
J1	3	1463.7665	1464.7738	VYKPHypASHypSPVYK
A3	4	1504.7316	1505.7389	GVPGADIFYEANPR
A1	5	2327.1340	2328.1413	FPHVANSGEWPDLVYVVNER

Hyp, hydroxyproline; masses listed are monoisotopic

Table 2. Summary of Instrument Configuration and Ionization Mode Utilization

Mass Analyzer[1]	Number of laboratories[2]	Average score[3]
Single instrument used		
q/TOF	13	38.5
Ion trap	10	24.1
TTOF	9	43.2
One or more instruments used[4]		
q/TOF +	22	41.7
Ion trap +	18	31.1
TTOF +	13	46.4
Ionization mode		
ES	24	34.9
MALDI	13	43.8
ES and MALDI	6	35.0

[1] q/TOF, quadrupole time-of-flight; Ion trap, 3-D or linear trap; TTOF, tandem time-of-flight

[2] Four laboratories did not report the ionization method used.

[3] Average total score for all peptides analyzed by the indicated instrument or ionization mode

[4] This category represents each instance of the use of the indicated instrument. A number of laboratories reported use of more than one mass spectrometer to generate sequence information; however, details were not always provided about which specific instruments were used for each sequence analysis. For this table, if a specific instrument was listed, it was included in the appropriate category.

RESULTS

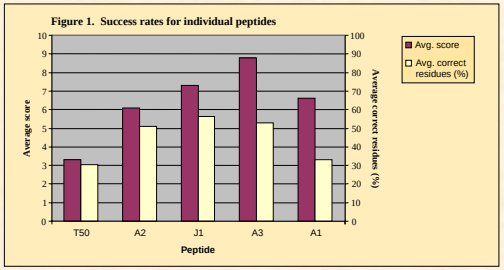
Sequence data were submitted by 48 laboratories. A summary of the study results, organized according to instrument configuration and ionization method, is shown in Table 2. A compilation of all results received is shown in Table 3. The following approaches were used: MS alone (42); Edman degradation (1); Edman degradation plus MS (5).

The great majority of laboratories reported the correct nominal peptide masses; peptide A2 was often found to contain an oxidized Met.
Differences in sample preparation and use of derivatization prior to analysis did not seem to influence the success rate for sequencing, although one group used a variety of derivatization strategies and obtained the correct sequence for 4 of the 5 peptides.
Static nanospray worked as well as on-line prefractionation by capillary HPLC.

Laboratories using a TTOF generally had a slightly higher success rate in obtaining the correct sequences for these peptides. These instruments typically use MALDI ionization; for this study it was not possible to assess the relative importance of ionization mode *versus* instrument type as related to the TTOF results. In addition, the scores for laboratories reporting use of both an ion trap and another type of instrument were notably higher than those using a trap alone. But, it is important to remember that there is a wide range of capabilities and levels of expertise among the participating laboratories, and the total number of responses is not very large.

Manual verification was used by all laboratories; software alone did not appear to be sufficient to provide complete sequences.
The success rates for sequencing the individual peptides varied (Fig. 1). This is most likely due to differences in the sequences. The internal Lys residues combined with the multiple Leu and Ile (scored as 0.5 if not distinguished) undoubtedly contributed to the low scores for peptide T50. Peptide A1 was the longest and, therefore, expected to be more difficult.

Figure 1. Success rates for individual peptides



DISCUSSION

The purpose of this study was to evaluate the capabilities of core laboratories to determine the sequences of peptides not found in any published database. Overall, the results show that this is an area that needs improvement in many core laboratories. A sufficient amount of each of the peptides was supplied such that sample quantity should not have been a limitation (although solubility issues might have caused problems for sequencing of peptide A1). Peptides T50 and A1 were the most difficult, probably due to specific sequence features of those peptides.
In general, laboratories that reported using more than one type of instrument did slightly better than those that only used a single instrument. It is possible that facilities with multiple instruments might have a larger staff with more overall expertise. Too few cases in which Edman sequencing was used were reported to draw any conclusions. However, quantity limitations and time constraints make it generally not possible to separate the peptides in a digest sufficiently for Edman analysis.

Although software is available that is designed to perform *de novo* sequencing, these programs do not as yet appear to be totally reliable. The peptides used in this study were by design not naturally occurring sequences. In many "real" cases, a partial sequence obtained by mass spectrometry, even with errors, can be linked to a protein by a BLAST search. But that would require that a protein of sufficient homology be present in a published database. While that strategy would not be successful for the synthetic peptides provided in this study, it should be routinely considered.
It is clear that manual interpretation was necessary for the peptides analyzed in this study. Commercially available instruments can usually provide sufficient information to determine the sequences of most unknown peptides. However, it is critically important not only to acquire the spectra with the requisite mass accuracy and resolution, but also to be skilled in data interpretation. For example, there are two Hyp residues in peptide J1. The residue mass of Hyp (113.04768) is 36.4-mmu lower than I/L (113.08406). Using some commercial instruments, it is possible to measure CID fragment masses with sufficient accuracy to distinguish between these residues.

Finally, expertise in *de novo* sequencing is clearly essential, regardless of whether the data are acquired by mass spectrometry or Edman analysis or both. Whereas proteins that are present in a published database can be identified on a routine basis by scientists who are not experts in interpretation of mass spectra, the same cannot be said for proteins for which sequences are not included in any database. The results of this study provide excellent justification for core laboratories to have not only state-of-the-art instrumentation but also personnel with expertise in instrument operation and data analysis.

CONCLUSIONS

- Most core laboratories need to improve their capabilities for *de novo* sequencing. The average success rate in this study was relatively low. Note that this study addresses issues that are very different from identifying a protein that is in a database.
- MALDI ionization and TTOF instrumentation appeared to be more successful than the alternatives, but the numbers of participating laboratories are too small to make this finding statistically reliable.
- No sample preparation or derivatization strategy was notably more successful.
- Laboratories that used more than one type of instrument were slightly more successful than those that only used a single type of instrument.
- Software for *de novo* sequencing alone was not sufficient for successful sequence analysis of the test peptides.
- Expertise in MS data acquisition and manual interpretation is essential for success.

ACKNOWLEDGMENTS

We thank David S. King of the HHMI Mass Spectrometry Laboratory at the University of California, Berkeley, for synthesis and purification of peptides A1, A2, and A3, Joe Leykam at the Macromolecular Structure Facility at Michigan State University for synthesizing peptide J1 and for the amino acid analyses, Ron Beavis and Janet Brostowin at the NYU Protein Chemistry Laboratory for the synthesis of peptide T50, Vivek Shetty, Chongfeng Xu and Yun Lu of the NYU Protein Analysis Facility for mass spectrometry analysis of the samples, Dawn Maynard of the NIMH at the National Institutes of Health for mailing and receiving correspondence and for ensuring that the participants remained anonymous, and Debra Diana of the NYU Skirball Institute of Biomolecular Medicine for receiving confirmatory data

Table 3. Results compilation

Identifier		Total score	Score	T50	Score	A2	Score	J1	Score	A3	Score	A1	Ionization method	Instrument type
		70	**11.0**	**LGAILKKLIPK**	**12.0**	**AYTFNMGQHSLK**	**13.0**	**VYKPHypASHypSPVYK**	**14.0**	**GVPGADIFYEANPR**	**20.0**	**FPHVANSGEWPDLVYVVNER**		
13579A		66.0	7.0	KLILQKLIPK	12.0	AYTFNMGQHSLK	13.0	VYKPHypASHypSPVYK	14.0	GVPGADIFYEANPR	20.0	FPHVANSGEWPDLVYVVNER	MALDI	TTOF
72079		64.0	8.5	(L/I)GA(L/I)(L/I)KK(L/I)(L/I)PK	11.5	AYTFN MoxGQHS(L/I)K	13.0	VYKPHypASHypSPVYK	13.5	GVPGAD(L/I)FYEANPR	17.5	FPHVANSWWPD(L/I)VYVVNER	ES	q/TOF
715		64.0	9.0	LGAILKIQIPK	12.0	AYTFNMGQHSLK	13.0	VYKPHypASHypSPVYK	14.0	GPVGADIFYEANPR	16.0	FPHVANSWWPD(I/L)VYVV(K/Q)DR	ES, E	q/TOF
20107	V	63.0	8.5	(I/L)GA(I/L)(I/L)KK(I/L)(I/L)PK	11.5	AYTFNMGQHS(I/L)K	13.0	VYKPHypASHypSPVYK	13.5	GPVQD(I/L)FYEANPR	16.5	(Mh(i/l))VANSGEWPD(I/L)VYVVNER	ES	LIT-FT, q/TOF
26019		62.3	7.5	(?)GA(I/L)(I/L)Q(K)K(I/L)(I/L)PK	11.5	AYTFNMGQHS(I/L)K	12.3	VYKP(I/L/Hyp)ASHypSPVYK	12.0	GVPGAD(?)FYEAGGPR	19.0	FPHVANSGEWPD(?)VYVVNER	ES	q/TOF
65214		61.5	6.0	KHyp(I/L)(L/I)KK(I/L)(I/L)PK	11.5	AYTFNMoxGQHS(L/I)K	13.0	VYKPHypASHypSPVYK	11.5	RPGAD(L/I)FYEANPR	19.5	FPHVANSGEWPD(L/I)VYVVNER	MALDI	q/TOF, TTOF, RTOF
10266	V	59.0	8.5	(L/I)GA(L/I)(L/I)KK(L/I)(L/I)PK	11.5	(AY)TFNMoxGQHS(L/I)K	11.0	VYKP(L/I)AS(L/I)SPVYK	8.5	rpQD(L/I)FYEANPR	19.5	FPHVANSGEWPD(I/L)VYVVNER	ES	LIT, LIT-FT
46011		58.0	7.0	LGALLKKLLPK	12.0	AYTFNMGQHSLK	5.0	VYKP(ps)S(twea/cp's)(qc)qK	14.0	GVPGADIFYEANPR	20.0	FPHVANSGEWPDLVYVVNER	MALDI	RTOF, TTOF (PSD)
dsu	V	56.0	4.0	klIkqllpk	9.0	YATFNMGQHSLK	13.0	VYKPHypASHypSPVYK	12.0	GVPGADLFYEAGGPR	18.0	FPHVANSGEWPDLVYVVDQR	MALDI	TTOF
12800		52.5	4.5	vvR(I/L)KKHypHypPK	8.0	(AY)TFEGMLHSLK	11.0	vykplasispvyk	13.0	GVPGADLFYEANPR	16.0	FPHVANSTTPDLVYVVSG(GE)R	MALDI	TTOF
78364		52.0	7.5	(I/L)GA(I/L)(I/L)(K/Q)(K/Q)(I/L)(I/L)PK	9.5	(AY)TFNMox(K/Q G)HS(I/L)K	7.0	FD(K/Q)P(I/L)AS(I/L)SPVYK	10.5	RPGAD(I/L)(F/Mox)YEANPR	17.5	(I/L)(M)HVANSGEWPD(I/L)VYVVNER	ES	LIT
13579B	V	52.0	8.0	(I/L)GA(I/L)(I/L)KK(Hyp)(I/L)PK	11.5	AYTFNMGQHS(I/L)K	12.0	VYKP(Hyp)AS(I/L)SPVYK	13.5	GPVGAD(I/L)FYEANPR	7.0	F[91.9][212.1]VYVV[243.1]	ES, MALDI	3DIT, RTOF, q/TOF
55000		51.5	6.0	KHyp(I/L)(I/L)KK(I/L)(I/L)PK	9.5	YATFNMGQHS(I/L)K	5.0	K[134]KP(I/L)AS(I/L)SHS(CP)K	11.5	VGPGAD(I/L)FYEANPR	18.5	FPHVANSGEADPD(I/L)VYVVNER	ES, MALDI	LIT, q/TOF
51565		51.0	4.0	KLLKHypKLLPK	9.0	HPTFNMGQHSHypK	9.0	VYQPLASLSPVYK	11.0	RPGADLFYEANPR	18.0	FPHVANSWWPDLVYVVNER	MALDI	TTOF, PSD
30109		48.8	5.0	Q(I/L/Hyp)(I/L/Hyp)LKK(I/L/Hyp)(I/L/Hyp)PK	9.5	PHTFNMGQHS(I/l)K	12.0	VYKPIAHypSPVYK	12.0	RPGADIFYEANPR	10.3	[568.3]PSWPD(I/L/Hyp)VYVVNER	MALDI	TTOF, q/TOF
11010		48.3	7.0	L(?)AILKKLIDL	10.0	AYTFNMGQH(L/I)SK	13.0	VYKPHypASHypSPVYK	12.0	(VG)PGADIFYEANPR	5.3	[235.19]fv[214.05]pfw[212.15](I/L/Hyp)VYVV[243.15]	ES, MALDI, E	3DIT, q/TOF
47223		44.5	6.0	ga(I/l)(I/l)psgag(I/l)(I/l)pk	1.0	ag(I/l)spqvsm(I/l)hpck	7.0	vy(k/q)p(II)apcpsvyk	14.0	GVPGADIFYEANPR	16.5	FPHvanswadpd(I/l)vyvvnER	MALDI	TTOF (PSD)
4318	V	42.0	2.0	KL(L/I)K(I/L)K(I/L)K(I/L)HypPK	10.0	YATFNMGQHSLK	8.0	VYKPHypASLSDFPK	4.0	GVDNAEYFLAGPRR	18.0	FPHVANSGEWPDLVYVVDQR	MALDI	TTOF
51952		41.0	2.0	(K/Q)(I/L)(I/L)(I/L)(K/Q)(K/Q)(I/L)(I/L)PK	11.0	AYTFNMoxG(K/Q)HS(I/L)K	12.0	VY(Q/K)PHypAS(I/Hyp)SPVYK	7.0	rP(K/Q)DLFyEAnpR	9.0	[938.57]fWpdLVYVV[243.14]R	ES	q/TOF
99999		41.0	5.0	(I/L)(Q/K)(I/L)(I/L)(QK)(QK)(I/L)(I/L)PK	11.0	AYTFNMoxG(Q/K)HS(I/L)K	11.0	VYKP(I/L)AS(I/L)SPVYK	11.0	GVPGAD(I/L)FYcgPGPR	3.0	(CS)(I/L)NVVYV(I/L)DP[1110.5]	ES	q/TOF
73108		40.0	2.0	KLLKLGALLPK	11.0	AYTFNMGKHSLK	11.0	VYKPASLSPVYK	7.0	GVPSLRFYEPGKR	9.0	PDLVYGFFWPDLVYVVGWR	ES	q/TOF
17999		40.0	6.0	(I/L)GA(I/L)(I/L)(Q/K)(I/L)AG(I/L/Hyp)PK	9.5	YATFNMG(Q/K)HSLK	11.0	VYK P(I/L)AS(I/L)SVPYK	13.5	GVPGAD(I/L)FYEANPR			ES	q/TOF
98166		38.0	7.5	(I/L)GA(I/L)(I/L)(K/Q)(K/Q)(I/L)(I/L)PK	10.5	AYTFNFG(K/Q)HSHypK	9.5	VY(K/Q)P(I/L)ASs(iI)PVYK	10.5	RPGAD(I/L)FYEAggPR	0.0	TFNFg(k/q)HSHypK	ES	3DIT, q/TOF
27406		38.0	7.0	(I/L)GA(I/L)(Q/K)(Q/K)(I/L)(I/L)PK	11.0	AYTFNMG(K/Q)HS(L/I)K	8.5	(YV)(K/Q)P(I/L)AS[279.12]VYK	11.5	VGPGAD(I/L)FYEANPR	0.0	AYTFNMoxG(Q/K)HS(L/I)K	ES	q/TOF
91741		34.5	3.5	(Q/K)(I/L)RVVK(I/L)(I/L)P(Q/K)	8.5	HPTFNMG(Q/K)HS(I/L)(Q/K)	3.5	PHGVPIASPCPVY(Q/K)	14.0	GVPGADIFYEANPR	5.0	FNFASEGWWLVLVYVVRDK	MALDI	TTOF
91573		34.0	1.0	TFNMoxGQHS(I/L)K	11.5	AYTFNMoxGQHS(II,)K	10.0	VYQP(I/L)AS(I/L)SpvYK	11.5	[158]PGAD(I/L)FYEANPR	0.0	T(I/L)(I/L)VNGVMYF[400]	ES	q/TOF, 3DIT
51583	V	32.0	2.0	KLLLKKLHypPK	1.0	PHTFNMGKHSHypK	5.0	AYRPIASISPVYK	12.0	RPGADIFYEANPR	12.0	EDNVANSVSVSHypVHypVYVVNER	MALDI	TTOF
70091		31.0	6.0	LGAIK[467.2]PK	1.0	[221.0]PA[427.0]NFTSK	10.0	VYKPHypASHypGPKYK	14.0	GPVGADIFYEANPR	0.0	[678.2][424.1]wegs(I/lHyp)av[381.3]	ES, E	LIT
19351		30.0			6.0	AYTFNM	12.0	VYKPLASHypSPVYK	12.0	(VG)PGADIFYEANPR			ES, E	q/TOF, LIT-FT
27974		29.5	0.0	[242.6](I/L)(Q/K)(I/L)(Q/K)(I/L)[356.5]	12.0	AYTFNMGQHSLK	6.0	VYKP(I/L)AS(I/L)SHDPR	9.5	RP(K/Q)D(I/L)FYEANPR	2.0	(?)VVYV(I/L)DPW(?)	ES	q/TOF
17017		26.0	4.5	(Q/K)Hyp(I/L)(Q/K)(Q/K)(I/L)(I/L)PK-NH_2			13.0	vVYKPHypASHypSPVYK	8.5	RPQD(I/L)FYEANPR			ES, MALDI, E	3DIT, PSD
12144		25.0	0.0	aygplvpvslppr	2.0	agplascppvyk	11.0	VYKPLASLSPVYK	12.0	gvpgadlfyeaggpr			ES	q/TOF
32466		22.0	1.0	KD(I/l)(I/l)qkasyK	11.5	AYTFNMGQHS(I/L)K	1.0	cmTFNkgftfsLK	8.5	RPQD(I/L)FYEANPR			ES, MALDI	3DIT
78544		21.3	0.0	VY(Q/K)KAPS(L/I)SAPYR	0.0	[235]T(Mox/F)(114)(Mox/F)[185]	9.8	VY(Q/K)P(I/L/Hyp)AS[200]PVYK	11.5	RPGAD(I/L)FYEANPR			ES	LIT
1467		19.5	5.0	IgalKGALI/(I/L)PK	1.0	(L/I)(Q/K)SHPSMNFTSK	1.0	[649]vasapqdk	4.0	RPQD(I/L)FYEANPR	8.5	RPQD(I/L)FYEANPR	ES	3DIT
52104		19.5	1.0	(I/L)PASHypSPVYK			12.5	VY(K/Q)PHypASHypSPVYK	6.0	PQDLFYEAGGPR + neutral loss of 157			ES, MALDI	q/TOF
80053		19.5									19.5	FPHVANSGEWPD(I/L)VYVVNER		
54321		18.5					8.0	VYKP(L/I)ASHCRYK, +Cys(ox) [+16,32,48]	10.5	RPGAD(L/I)FYEAGGPR			ES	q/TOF
85035	V	18.0	5.0	LGA(?)(?)QQ(?)(?)PK	1.0	FSTF(?)FAVHTVK	3.0	YVKPMSSA(?)P(?)MK	4.0	RPGAD(?)FFNHGK	5.0	[1449.6094]VYVVGWR	ES	q/TOF
10523	V	12.0	0.0	[243.09](I/L)(I/L)[126]DK[355.36]	0.0	[145.09]DSH(I/L)(I/L)GQF[336.44]	6.0	[263.13]KP(I/L)ASH[422.82]K	4.0	GVDRNAGSYFA(I/L)PR	2.0	F[2005.2]R	MALDI	RTOF, IT/TOF
87458		10.5	0.0	(RD)(P Q/K)L(F/Mox)YEAN[315.01]	4.5	(YA/FS/HP/MC)T(Mox/F)NMG(Q/K)H(EA/TV/CP)	0.0	[262.32](Q/K)P(LA/PS)S(TV/CD/LS/EA)[487.19	6.0	(Q/K P)DL(F/Mox)YEANPR	0.0	(YA/FS/HP/MC)(F/Mox)AYYVLDPW[920.54]	MALDI	TTOF
1605		8.8	0.0	(K/Q)Hyp(I/L)(I/L)G(K/Q)AHyp(I/L)PK	6.5	YATFNMNAS(I/L)K	1.0	VQNNMoxYEANPR	1.3	ED(I/L/Hyp)(K/Q)TNHPK	0.0	MoxDQPHypASAEDEDK	ES	LIT
12345		5.0	0.0	V(?)KPLAKHypIPVN	5.0	AYTFHypMIFH(?)Lykr	0.0	L(?)AIAKSLSEA	0.0	SPLVNDGQE(?)K			E	
11747		1.0			1.0	FSTFNMSYASMK							ES	LIT
7974		0.0	0.0	KN(I/L/Hyp)									ES	3DIT
49495		0.0							0.0	(VY)(Q/K)PSP(I/L)S(PV)Y(K/Q)				
47551		0.0	0.0	VYKPHypASHypSPVYK(K)										
11089		0.0												

A "V" shown after the identifier indicates that the participating laboratory is an instrument vendor.

The peptide score represents the sum of consecutive correct residues as follows: score = xC + yN + zM, where the number of consecutive correct residues starting at the *C*-terminus is xC; at the *N*-terminus, yN; and in the middle, zM. Lack of differentiation between isobaric or nearly isobaric residues was scored as follows: Ile/Leu, 0.5; Gln/Lys, 0.5 Gln/Lys/Hyp, 0.3.

Abbreviations: 3DIT, 3-dimensional ion trap; E, Edman; ECD, electron capture dissociation; ES, electrospray; Hyp, hydroxyproline; LIT, IT-TOF, ion trap/time-of-flight; linear ion trap; LIT-FT, linear ion trap/Fourier transform hybrid; MALDI, matrix-assisted laser desorption ionization; Mr, relative molecular mass; Mox, oxidized Met; PSD, post-source decay; q/TOF, quadrupole/time-of-flight; RTOF, reflectron time-of-flight; TTOF, tandem time-of-flight. The correct sequence is shown on the first results line. All methods and instruments used by a laboratory are listed together; in a few cases, different methods/instruments were used for different peptides. The collision energy used depended on the instrument type and is not specified in the table. Some groups also used PSD and one used ECD, as noted. Additional details can be found at **http://www.abrf.org/index.cfm/group.show/Proteomics.34.htm**.