



SEQUENCE 4.0

de novo peptide sequencing software

**Protein identification using *de novo* sequencing
of proteolytic peptides by MALDI-MS**
Application Note

Protein identification using de novo sequencing of proteolytic peptides by MALDI-MS
SEQUIT! APPLICATION NOTE

R. Demine¹ and P. Walden²

¹ Proteome Factory AG

² Department of Dermatology and Allergy, Charité – University Medicine Berlin, Humboldt University Berlin, D-10098 Berlin, Germany

Sequit! - software for *de novo* peptide sequencing by tandem mass spectrometry

© 2005, Proteome Factory AG, Berlin, Germany - www.proteomefactory.com

Based on an Invention by the Charité – Universitätsmedizin Berlin, Germany
Licensed by IPAL, Germany

PURPOSE

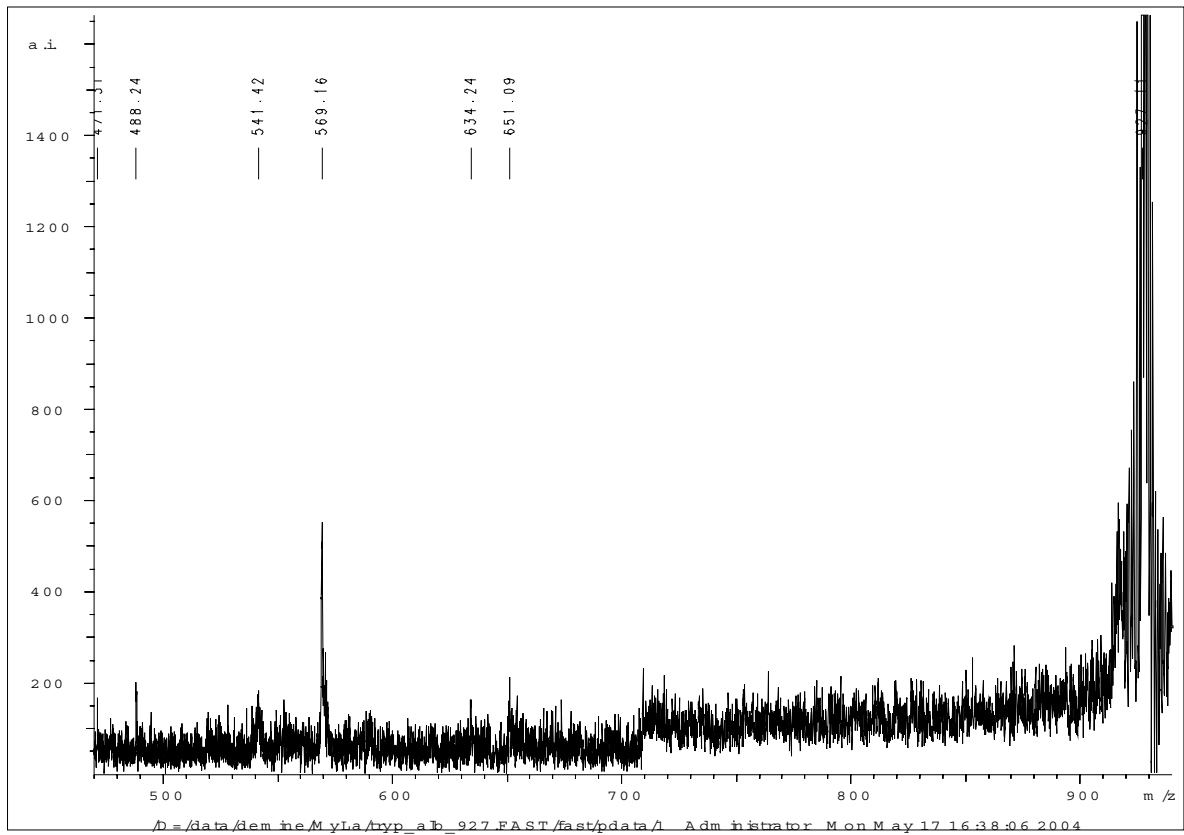
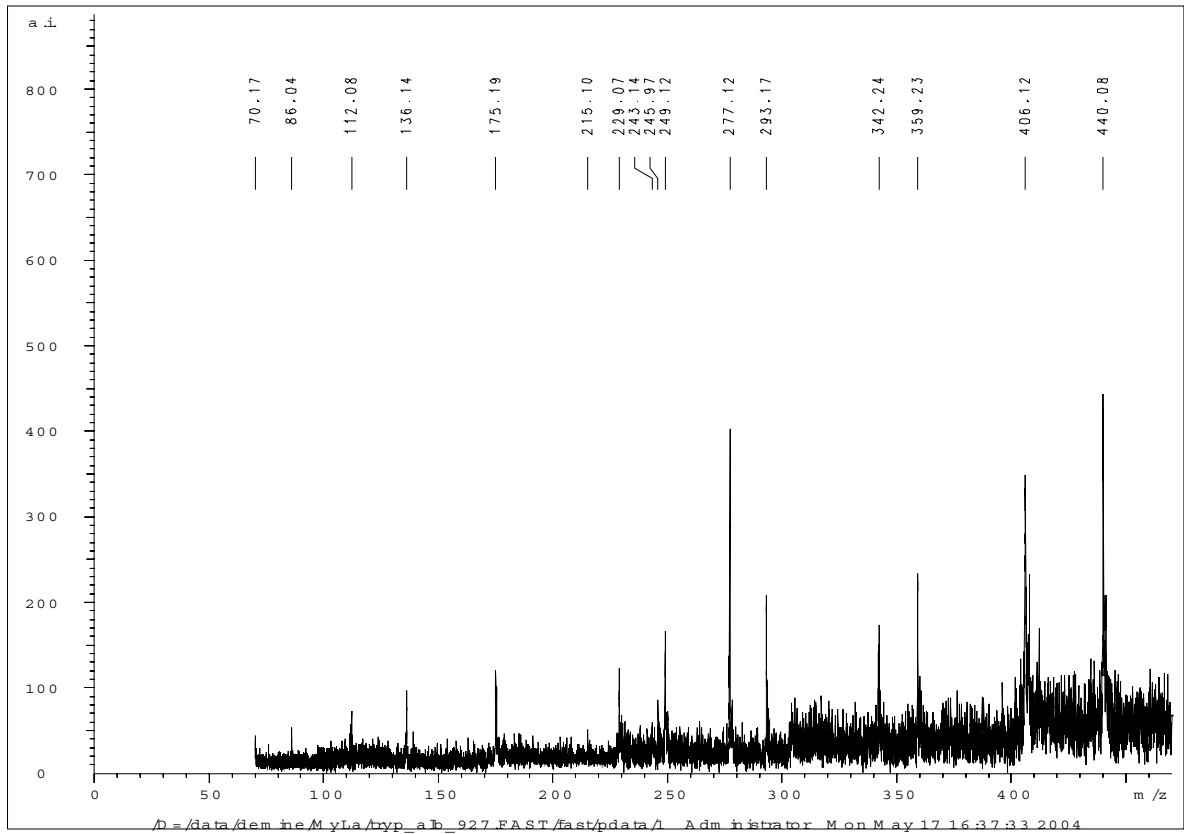
The major strategy in protein identification is to match mass spectrometric data of proteolytic protein fragments (PMF, Peptide Mass Fingerprint) to database entries. However, in the case when no significant match is obtained because of low numbers of detectable proteolytic fragments or when unknown proteins are analyzed, alternative protein identification strategies are needed.

EXPERIMENTAL PROCEDURE

Human proteins were separated by SDS PAGE. PMF of the trypsin digest of an SDS PAGE gel band was measured by MALDI-MS. 927.52 Da and 960.54 Da peptides of PMF were sequenced by MALDI-*Post Source Decay* (PSD) (1). MALDI-PSD measurements were performed with a Bruker MALDI-TOF Reflex II mass spectrometer with α -HCCA as matrix. The accuracy of peptide mass measurement was 0.2 Da, and that of peptide fragment mass measurement was 0.5 Da.

Note:

Since modern MALDI-TOF/TOF fragmentation techniques and MALDI-PSD are very similar in peptide fragmentation behavior, no further Sequit! optimization is required for MALDI-TOF/TOF MS/MS data.



MALDI-PSD spectra of the 927.52 Da peptide. (MALDI-PSD spectra of the 960.54 Da peptide is not shown.)

Sequit! input files were generated from the MALDI-PSD data:

927.sqi

```
Sample = 927
[M+H]+ = 927.52
Peptide mass tolerance = 0.1
Fragment mass tolerance = 0.5
```

Mass Intensity

```
70.17 0.0033
86.04 0.0040
112.08 0.0056
136.14 0.0073
175.19 0.0092
215.10 0.0039
229.07 0.0093
243.14 0.0045
245.97 0.0065
249.12 0.0126
277.12 0.0306
293.17 0.0159
342.24 0.0132
359.23 0.0177
406.12 0.0264
440.08 0.0334
471.31 0.0127
488.24 0.0155
541.42 0.0140
569.16 0.0415
634.24 0.0123
651.09 0.0160
927.11 1.0056
```

960.sqi

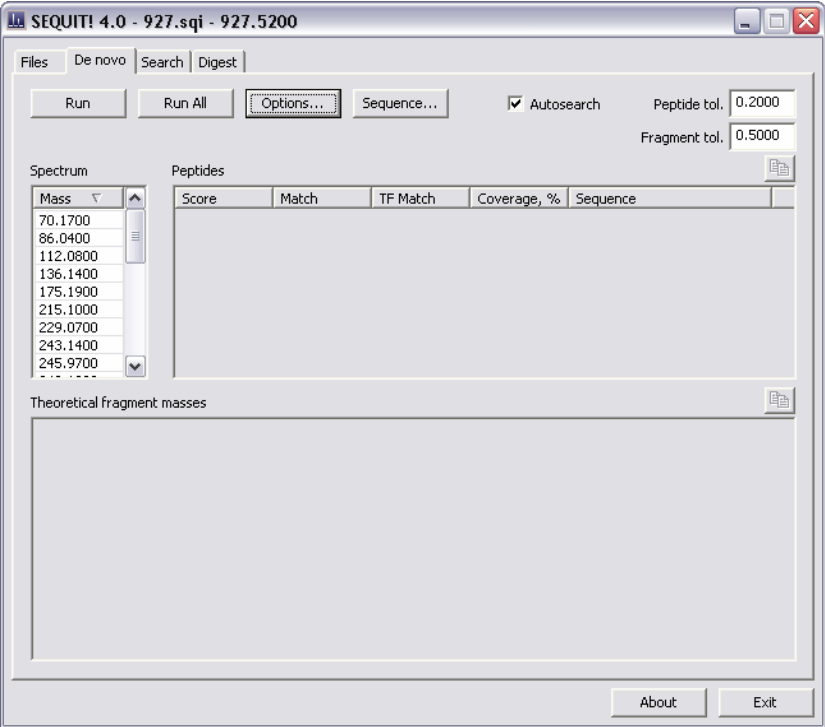
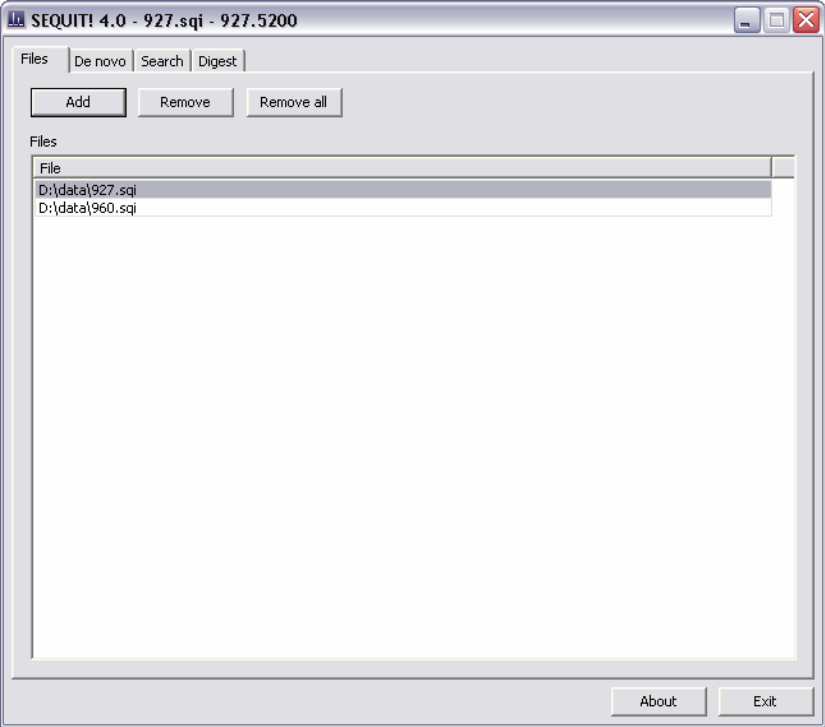
```
Sample = 960
[M+H]+ = 960.54
Peptide mass tolerance = 0.1
Fragment mass tolerance = 0.5
```

Mass Intensity

```
112.06 0.0021
120.00 0.0035
165.05 0.0056
175.01 0.0209
185.11 0.0038
186.04 0.0061
243.37 0.0047
248.11 0.0070
257.16 0.0100
259.14 0.0065
274.15 0.0092
276.05 0.0316
293.11 0.0066
298.14 0.0072
299.22 0.0135
314.40 0.0073
342.09 0.0115
344.94 0.0084
370.28 0.0123
387.24 0.0097
390.04 0.0386
410.15 0.0111
412.21 0.0095
415.96 0.0091
426.95 0.0115
433.04 0.0143
444.02 0.0148
461.08 0.0909
483.35 0.0128
500.24 0.0115
512.50 0.0117
529.08 0.0199
546.02 0.0220
574.17 0.0225
668.12 0.0137
791.29 0.0170
810.55 0.0147
```

The [M+H]⁺ values 927.52 Da and 960.54 Da were obtained from peptide mass measurement that is more accurate than MALDI-PSD measurement.

Both input files were loaded to Sequit! and processed as batch performing successive *de novo* sequencing and database search steps for each file.



RESULTS

Four sequences with the same C-terminal sequence tag YELAR were computed the highest score of 2.85 for 927.52 Da peptide. Combinations of the amino acids E and F or L and Y were proposed for positions 1 and 2. Such results with a series of similar sequences are typically obtained for sequence computation using spectra with incomplete fragment series and aberrant peaks. 10 best scored proposed sequences and their L/I variants were used for database search.

SEQUIT! 4.0 - 927.sqi - 927.5200

Files De novo Search Digest

Run Run All Options... Sequence... Autosearch Peptide tol. 0.2000
Fragment tol. 0.5000

Mass	Score	Match	TF Match	Coverage, %	Sequence
70.1700	2.85	20	14	83.33	EFVELAR
86.0400	2.85	20	14	83.33	FEVELAR
112.0800	2.85	20	14	83.33	LYVELAR
136.1400	2.85	20	14	83.33	YLYELAR
175.1900	2.57	18	12	75.00	YLYEPR
215.1000	2.57	18	12	75.00	YLYEPR
229.0700	2.42	17	12	70.83	EFEVELAR
243.1400	2.42	17	12	70.83	LYYEPR
245.9700	2.42	17	12	70.83	LYYEPR

Ion	1	2	3	4	5	6	7
a-18							
a-17		232.1332	395.1965	524.2391	637.3232		
a		249.1597	412.2230	541.2656	654.3497		
b-18							
b-17		260.1281	423.1914	552.2340	665.3181	736.3552	
b		277.1546	440.2179	569.2605	682.3446	753.3817	
b+18						771.3923	
y	Y						
y-17		764.4302	651.3461	488.2828	359.2402	246.1561	175.1190
y-18		747.4037	634.3196	471.2563	342.2137	229.1296	
immonium		136.0757	86.0965	136.0757	102.0550	86.0965	44.0495
							129.1135

About Exit

YLYEIAR was identified by a subsequent database search as peptides derived from the serum albumin. No matches were found for other proposed sequences.

SEQUIT! 4.0 - 927.sqi - 927.5200

Files De novo Search Digest

Search Options... Organism any

Query	Score	Letters	Identities	Positives	BLAST Score	E
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30
YLYEIAR	2.85	7	7/7	7/7	27.40	2.30

Query details

YLYEIAR 2.85 7 7/7 7/7 27.40 2.30

>gi|23307793|gb|AAN17825.1| serum albumin [Homo sapiens]
Length = 609

Score = 27.4 bits (57), Expect = 2.3
Identities = 7/7 (100%), Positives = 7/7 (100%)

Query: 1 YLYEIAR 7
YLYEIAR
Sbjct: 162 YLYEIAR 168

About Exit

960.54 Da peptide was identified as FQNALLVR from the same protein.

SEQUIT! 4.0 - 960.sqi - 960.5400

Files De novo Search Digest

Run Run All Options... Sequence... Autosearch Peptide tol. 0.2000
Fragment tol. 0.5000

Mass	Score	Match	TF Match	Coverage, %	Sequence
112.0600	3.75	30	20	76.92	FKNALLVR
120.0000	3.75	30	20	76.92	FQNALLVR
165.0500	3.37	27	20	69.23	KFNALLVR
175.0100	3.37	27	20	69.23	QFNALLVR
185.1100	3.33	30	20	76.92	FKGGALLVR
186.0400	3.33	30	20	76.92	FQGGALLVR
243.3700	3.25	26	18	66.66	FKNALDPR
248.1100	3.25	26	18	66.66	FKNALVLR
257.1600	3.25	26	18	66.66	FQNALDPR

Theoretical fragment masses

Ion	1	2	3	4	5	6	7	8
a-18								
a-17		231.1128	345.1557	416.1928	529.2769	642.3610		
a		248.1393	362.1822	433.2193	546.3034	659.3875		
b-18								
b-17		259.1077	373.1506	444.1877	557.2718	670.3559	769.4243	
b		276.1342	390.1771	461.2142	574.2983	687.3824	786.4508	
b+18								804.4614
	F	Q	N	A	L	L	V	R
y		813.4942	685.4356	571.3927	500.3556	387.2715	274.1874	175.1190
y-17		796.4677	668.4091	554.3662	483.3291	370.2450	257.1609	
y-18								
immonium	120.0808	101.0710	87.0553	44.0495	86.0965	86.0965	72.0808	129.1135

About Exit

SEQUIT! 4.0 - 960.sqi - 960.5400

Files De novo Search Digest

Search Options... Organism any

Query	Score	Letters	Identities	Positives	BLAST Score	E
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10
FQNALLVR	3.75	8	8/8	8/8	28.60	1.10

Query details

```

FQNALLVR 3.75 8 8/8 8/8 28.60 1.10
>gi|23307793|gb|AA17825.1| serum albumin [Homo sapiens]
Length = 609

Score = 28.6 bits (60), Expect = 1.1
Identities = 8/8 (100%), Positives = 8/8 (100%)

Query: 1 FQNALLVR 8
      FQNALLVR
Sbjct: 427 FQNALLVR 434
    
```

About Exit

The assignment of YLYEIAR and FQNALLVR were verified by comparing the experimental MALDI-PSD spectrum with the theoretical peak list calculated for this peptide. Special considerations were taken with respect to masses that are characteristic for MALDI-PSD peptide fragmentation.

C-terminal arginine is indicated by 175.12 Da y1 ion. The y-Ion series is dominant which is characteristic for peptides with a C-terminal arginine.

Ion type	1	2	3	4	5	6	7
a-17		232.1332	395.1965	524.2391	637.3232		
a		249.1597	412.2230	541.2656	654.3497		
b-18							
b-17		260.1281	423.1914	552.2340	665.3181	736.3552	
b		277.1546	440.2179	569.2605	682.3446	753.3817	
b+18						771.3923	
	Y	L	Y	E	L	A	R
y		764.4302	651.3461	488.2828	359.2402	246.1561	175.1190
y-17		747.4037	634.3196	471.2563	342.2137	229.1296	
y-18							
immonium related	136.0757	86.0965	136.0757	102.0550	86.0965	44.0495	129.1135
related							112.0869
int2		277.1547	293.1132	243.1340	185.1285		
int3		406.1973	406.1973	314.1711			
int4		519.2814	477.2344				
int5		590.3185					

Ion type	1	2	3	4	5	6	7	8
a-17		231.1128	345.1557	416.1928	529.2769	642.3610		
a		248.1393	362.1822	433.2193	546.3034	659.3875		
b-18								
b-17		259.1077	373.1506	444.1877	557.2718	670.3559	769.4243	
b		276.1342	390.1771	461.2142	574.2983	687.3824	786.4508	
b+18							804.4614	
	F	Q	N	A	L	L	V	R
y		813.4942	685.4356	571.3927	500.3556	387.2715	274.1874	175.1190
y-17		796.4677	668.4091	554.3662	483.3291	370.2450	257.1609	
y-18								
immonium related	120.0808	101.0710	87.0553	44.0495	86.0965	86.0965	72.0808	129.1135
related		129.0659						112.0869
related		84.0444						
int2		243.1088	186.0873	185.1285	227.1755	213.1598		
int3		314.1459	299.1714	298.2126	326.2439			
int4		427.2300	412.2555	397.2810				
int5		540.3141	511.3239					
int6		639.3825						

Additional PMF analyses confirm the protein identification. Sequit! allows the calculation of theoretical digest of protein, which was identified by de novo sequencing (marked blue), and comparison of theoretical digest to the measured PMF. 16 from 32 peptides from measured PMF match to serum albumin (marked red). Resulting sequence coverage of 25% was calculated.

Additionally, Sequit! shows unmatched masses, which are possibly masses of peptides deviating from database protein sequence (mutated, alternatively spliced or modified peptides) or masses from other protein(s) or contaminants.

SEQUIT! 4.0 - 927.sqi - 927.5200

Files | De novo | Search | Digest

Open Clear PMF tolerance 0.1000 Missed cleavages 0

Sequence info

protein mass: 69,349 Da
sequence coverage: 0%
matched masses: 0 from 0
PMF tolerance: 0.1000
missed cleavages: 0

MKWWYTFISLLFLFSSAYSRGVFRDDAHKSEVAHRFKDLGEENFKALVLIIFAQYLQCCPFEDH-
VKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECF-
LQHKDDNPNLRLRPEVDVMCTAFHDNEETFLKKYLYEIAARRHPYFYAPELFFAKRYKAAFT-
ECCQAADKAAACLLPKLDELRLDEGKASSAKQRLKASLQKFGGERAFKAWAVARLSQRFPKAEF-

Unmatched masses

Mass

Tryptic protein fragments (mass range 500 - 3000 Da)

Theoretical mass	Measured mass	Error	Sequence
503.2936			LSQR
508.2514			FGER
509.3195			HKPK
517.2980			ADLAK
517.2981			EQLK
581.3042			HPEAK
591.3349			TTLEK
616.3776			LVAAASR
645.3566			LDELR
649.3338			CASLQK
658.3155			QEPER
673.3779			AWAVAR

About Exit

SEQUIT! 4.0 - 927.sqi - 927.5200

Files | De novo | Search | Digest

Open Clear PMF tolerance 0.2000 Missed cleavages 2

Sequence info

protein mass: 69,349 Da
sequence coverage: 25%
matched masses: 16 from 32
PMF tolerance: 0.2000
missed cleavages: 2

MKWWYTFISLLFLFSSAYSRGVFRDDAHKSEVAHRFKDLGEENFKALVLIIFAQYLQCCPFEDH-
VKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECF-
LQHKDDNPNLRLRPEVDVMCTAFHDNEETFLKKYLYEIAARRHPYFYAPELFFAKRYKAAFT-
ECCQAADKAAACLLPKLDELRLDEGKASSAKQRLKASLQKFGGERAFKAWAVARLSQRFPKAEF-
AEVSKLVDTLTKVHTECCHGDLLCADDRADLAKYICENQDSIS5SKLKECCPKLLEKSHCIAEV-

Unmatched masses

Mass

1093.4300
1213.5500
1331.8700
1430.6600
1452.6400
1483.7200
1489.6300
1583.1600
1591.7100

Tryptic protein fragments (mass range 500 - 3000 Da)

Theoretical mass	Measured mass	Error	Sequence
890.5129			LKASLQK
892.4370			DEGKASSAK
915.4176			CCKHPEAK
927.4934	927.4900	0.0034	YLYEIAAR
940.4482	940.4300	0.0182	DDNPNLPR
951.4418			DLGEENFK
960.5625	960.5400	0.0225	FQNALLR
973.5213			SEVAHRFK
983.5409	983.5000	0.0409	TYKTTLEK
988.5898			ASSAKQRLK
1000.6038			QTALVELVK
1002.5578			TPVSDRVTK

About Exit

CONCLUSION

De novo peptide sequencing of tryptic peptides with subsequent database search is a potent alternative to peptide mass fingerprint analyses for protein identification. At least one fragment from a proteolytic protein digest is required and suffices for protein identification.

REFERENCES

1. Demine R, Walden P. Sequit: Software for *de novo* peptide sequencing by matrix-assisted laser desorption/ionization post-source decay mass spectrometry. *Rapid Commun Mass Spectrom.* 2004; **18**: 907-913.
2. http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.shtml#Batch