



SEQUENCE 4.0

de novo peptide sequencing software

***De novo* sequencing of MHC class I-associated peptides**

Application Note

De novo sequencing of MHC class I-associated peptides

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

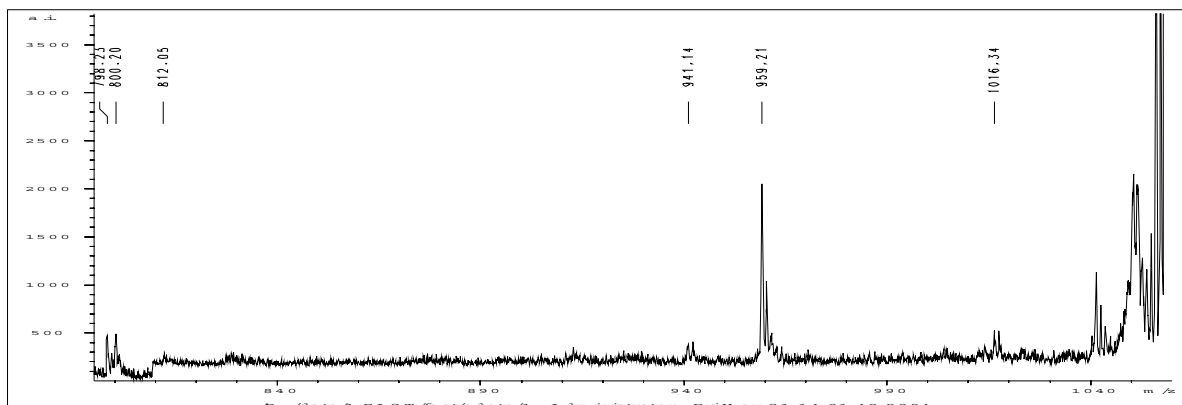
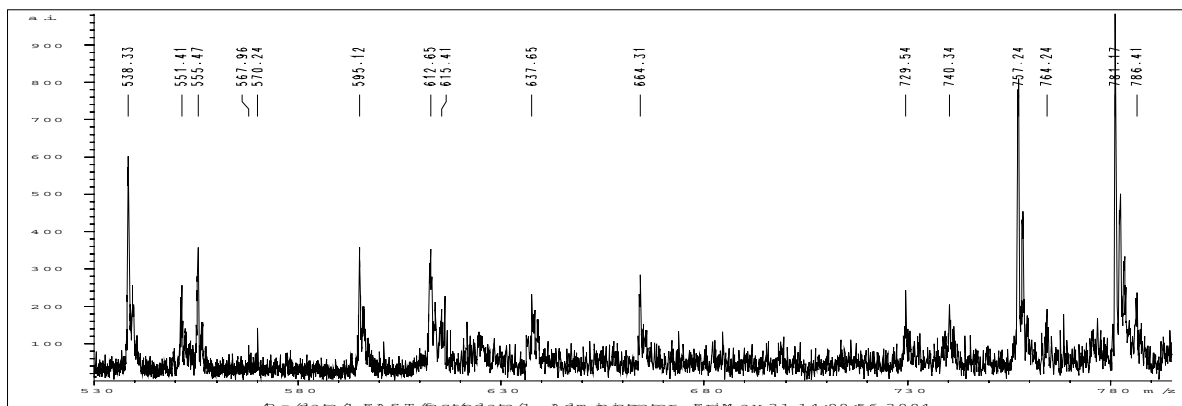
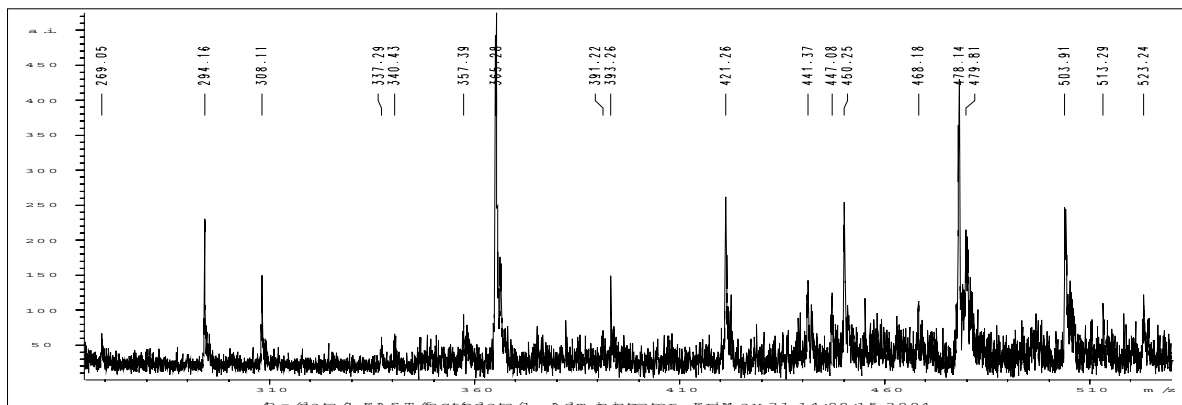
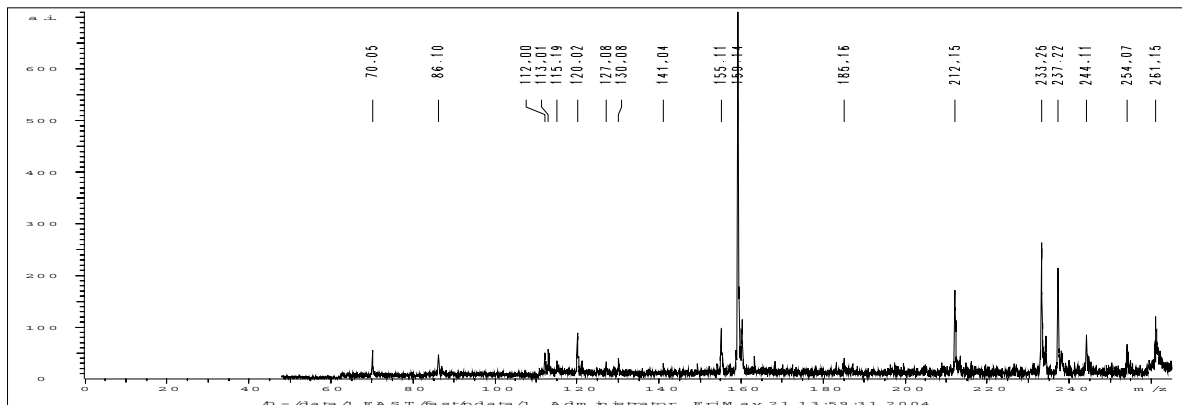
PURPOSE

The determination of the sequence of *major histocompatibility complex* (MHC) class I-associated peptides is important for understanding antigen processing and for the identification of tumor-associated or viral T cell epitopes. MALDI-PSD sequencing of a synthetic analogue of the HLA A2-restricted T cell epitope FLWGPRALV (1) is used here as an example.

EXPERIMENTAL PROCEDURE

MALDI-PSD measurements were performed with a Bruker MALDI-TOF Reflex IV mass spectrometer using α -HCCA as matrix. The accuracy of peptide mass measurement was 0.1 Da, and that of peptide fragment mass measurement was 0.4 Da.

MALDI-PSD spectra of the 1058.54 Da peptide.



The Sequit input file generated from the MALDI-PSD data:

```
Sample = 1
[M+H]+ = 1058.54
Peptide mass tolerance = 0.1
Fragment mass tolerance = 0.4

Mass Intensity
---
70.05 0.0036
86.10 0.0031
112.00 0.0034
113.01 0.0038
115.19 0.0023
120.02 0.0059
127.08 0.0022
130.08 0.0026
141.04 0.0020
155.11 0.0065
159.14 0.0470
185.16 0.0027
212.15 0.0114
233.25 0.0173
237.22 0.0141
244.11 0.0056
254.07 0.0044
261.15 0.0081
269.05 0.0044
294.16 0.0153
308.11 0.0098
337.29 0.0041
340.43 0.0043
357.39 0.0062
365.28 0.0349
391.22 0.0049
393.26 0.0097
421.26 0.0171
441.37 0.0096
447.08 0.0082
450.25 0.0167
468.18 0.0074
478.14 0.0283
479.81 0.0144
503.91 0.0166
513.29 0.0072
523.24 0.0080
538.33 0.0397
551.41 0.0169
555.47 0.0236
567.96 0.0063
570.24 0.0093
595.12 0.0235
612.65 0.0232
615.41 0.0128
637.65 0.0155
664.31 0.0189
729.54 0.0162
740.34 0.0135
757.24 0.0530
764.24 0.0127
781.17 0.0659
786.41 0.0155
798.23 0.0310
800.20 0.0324
812.05 0.0178
941.14 0.0252
959.21 0.1340
1016.34 0.0345
---
```

The $[M+H]^+$ value of 1058.54 Da was obtained from peptide mass measurement that is more accurate than MALDI-PSD measurement (1058.32 Da). $[M+H]^+_{\text{theor.}}$ is 1058.62 Da.

RESULTS

The computation of the input data with Sequit! yielded 10 best scored sequences (default value for MALDI-MS) that fit the MALDI-PSD spectrum to varying degrees.

The screenshot shows the SEQUIT! 4.0 interface with the following data:

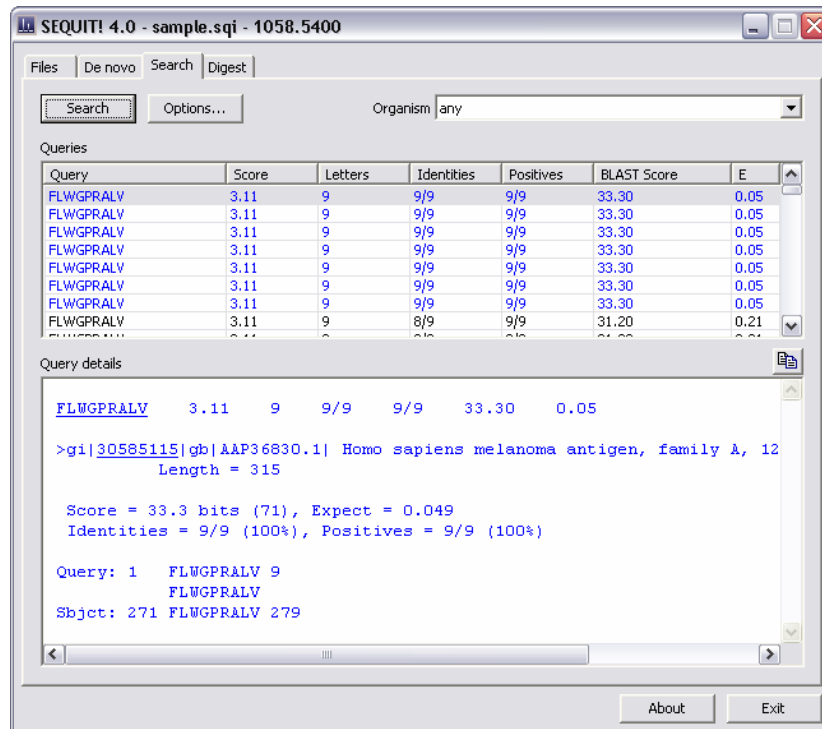
Mass	Score	Match	TF Match	Coverage, %	Sequence
70.0500	3.11	28	18	45.90	FLWGPRALV
86.1000	3.11	28	18	45.90	FLWGRPALV
112.0000	3.11	28	18	45.90	LFWGRPALV
113.0100	3.00	27	18	44.26	LFWGRPALV
115.1900	2.88	26	17	42.62	FLWGPRLAV
120.0200	2.88	26	17	42.62	FLWGRPSV
127.0800	2.88	26	17	42.62	FLWGRSPV
130.0800	2.88	26	17	42.62	FLWGRPLAV
141.0400	2.88	26	17	42.62	FLWGRPPSV

Ion	1	2	3	4	5	6	7	8	9
a-18									
a-17		216.1383	402.2176	459.2391	556.2919	712.3930	783.4301		
a		233.1648	419.2441	476.2656	573.3184	729.4195	800.4566		
b-18									
b-17		244.1332	430.2125	487.2340	584.2868	740.3879	811.4250	924.5091	
b		261.1597	447.2390	504.2605	601.3133	757.4144	828.4515	941.5356	
b+18								959.5462	
	F	L	W	G	P	R	A	L	V
y		911.5463	798.4622	612.3829	555.3614	458.3086	302.2075	231.1704	118.08
y-17		894.5198	781.4357	595.3564	538.3349	441.2821	285.1810	214.1439	
y-18									

Three proposed peptides with the highest score of 3.11 have amino acids WG on positions 3 and 4 and C-terminal sequence tag ALV. Combinations of the amino acids F, L and P, R were proposed for positions 1, 2 and 5, 6, respectively.

score	match	coverage	peptide
3.11	28	45.90	FLWGPRALV
3.11	28	45.90	FLWGRPALV
3.11	28	45.90	LFWGRPALV
3.00	27	44.26	LFWGRPALV
2.88	26	42.62	FLWGPRLAV
2.88	26	42.62	FLWGRPSV
2.88	26	42.62	FLWGRSPV
2.88	26	42.62	FLWGRPLAV
2.88	26	42.62	FLWGRPPSV
2.88	26	42.62	FLWGRPSV

This result with a series of similar sequences is an example for sequence computation based on MALDI-PSD spectra with incomplete fragment series and aberrant peaks. Nonetheless, the correct sequence is identified by subsequent database search. The Sequit! output file fasta.sqi contains 10 proposed sequences and their L/I variants. A batch search done for these sequences using a local BLAST version with the NCBI nr database (2) identified FLWGPRALV as a peptide derived from the Melanoma-associated antigens MAGE-3 or MAGE-12. No matches were found for the other 9 sequences and their L/I variants.



The correct assignment of FLWGPRALV was validated by comparing the experimental MALDI-PSD spectrum with the theoretical peak list calculated for this peptide with respect to the masses, which are specifically characteristic for MALDI-PSD peptide fragmentation. The b-Ion series is dominant due to the absence of a charged amino acid at the C-terminus. The b8+H₂O ion of 959.21 Da is indicative of a C-terminal V. The presence of F, L and W in the sequence is indicated by the immonium ions of 120, 86 and 159 Da, respectively. The 70 Da ion can be either an arginine-related ion or the proline immonium ion. The arginine-containing b6 and b7 fragments undergo carbonmonoxide decomposition producing the a6 and a7 fragments, respectively.

Ion type	1	2	3	4	5	6	7	8	9
a-18									
a-17		216.1383	402.2176	459.2391	556.2919	712.3930	783.4301		
a		233.1648	419.2441	476.2656	573.3184	729.4195	800.4566		
b-18									
b-17		244.1332	430.2125	487.2340	584.2868	740.3879	811.4250	924.5091	
b		261.1597	447.2390	504.2605	601.3133	757.4144	828.4515	941.5356	
b+18								959.5462	
	F	L	W	G	P	R	A	L	V
y		911.5463	798.4622	612.3829	555.3614	458.3086	302.2075	231.1704	118.0863
y-17		894.5198	781.4357	595.3564	538.3349	441.2821	285.1810	214.1439	
y-18									
immonium	120.0808	86.0965	159.0917	30.0339	70.0652	129.1135	44.0495	86.0965	72.0808
related						112.0869			
related									
int2		300.1707	244.1081	155.0816	254.1612	228.1455	185.1285		
int3		357.1922	341.1609	311.1827	325.1983	341.2296			
int4		454.2450	497.2620	382.2198	438.2824				
int5		610.3461	568.2991	495.3039					
int6		681.3832	681.3832						
int7		794.4673							

CONCLUSION

Sequence determination of the MHC class I-associated peptide presented here is possible from MALDI-PSD data using the Sequit! software combined with a database search.

REFERENCES

1. van der Bruggen P, Bastin J, Gajewski T, Coulie PG, Boel P, De Smet C, Traversari C, Townsend A, Boon T. A peptide encoded by human gene MAGE-3 and presented by HLA-A2 induces cytolytic T lymphocytes that recognize tumor cells expressing MAGE-3. *Eur J Immunol.* 1994; **24**: 3038-3043.
2. http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.shtml#Batch