Bioorganic & Medicinal Chemistry Letters, Vol.3, No.3, pp. 431-436, 1993
Printed in Great Britain

0960-894X/93 $6.00 + .00

NATURAL AND SYNTHETIC PEPTIDE POOLS: CHARACTERIZATION BY SEQUENCING AND ELECTROSPRAY MASS SPECTROMETRY

Stefan Stevanovic, Karl-Heinz Wiesmüller, Jörg Metzger, Annette G. Beck-Sickinger, Günther Jung*

Institut für Organische Chemie, Universität Tübingen, Auf der Morgenstelle 18,
D-7400 Tübingen, Germany

Abstract

Natural peptide pools consisting of octa- or nonapeptides and carrying defined anchor amino acids in at least two specific positions are bound to MHC-class I molecules. These isolated natural libraries and their synthetic analogues can be characterized by novel methods of multiple sequence analysis using automatic Edman degradation and electrospray mass spectrometry.

Introduction

The most sophisticated natural system known so far is the multimeric complex consisting of the MHC-class I molecule, in association with β_2-microglobulin presenting in its binding groove an oligopeptide antigen to the T-cell receptor of CTL's (cytotoxic T lymphocytes). Degradation of self and viral proteins within the cell creates thousands of different peptides which appear at the surface bound to MHC I in order to be recognized as foreign or self peptides. Peptides presented by one MHC allele have a common sequence motif [1]: they represent a *natural peptide library*. Viral peptides fitting this motif may be recognized by CTL's in attomolar concentrations. For the analysis of the extremely small amounts of isolated self peptide mixtures we have had to develope novel tools and evaluation methods for sequencing. These methods are also very valuable in characterizing *synthetic peptide libraries* [2-4]. The increasing interest in complex peptide mixtures for screening receptor ligands in hope finding new lead compounds for drug development will certainly require highly decisive and sensitive analytical methods. Herein we describe the application of pool-sequencing using Edman degradation and of triple quadrupole mass spectrometry using an electrospray interface.

Results and Discussion

Pool-sequencing of natural peptide libraries

The highest state of molecular diversity occurs in the mammalian immune system, where billions of different specificities are represented by one class of proteins like antibodies or T cell receptors. MHC-bound peptides, as counterparts of the T cell receptors, reflect this diversity. The bulk of MHC-bound self peptides is expected to exhibit the complete variety of interactions between peptides and one MHC-type. To investigate the rules allowing a distinct peptide to bind to a distinct MHC I molecule, mixtures of self peptides were isolated as peptide pools in the group of H.-G. Rammensee [1].

We analyzed these pools by Multiple Sequence Analysis using automated Edman degradation in a pulsed-liquid protein sequencer. The results (Table 1) revealed binding motifs that are allele-specific for the respective MHC I protein. In the example shown, the peptides that bound to the murine K^d-molecule shared the length of nine amino acids as well as a tyrosine in position 2 and a leucine or isoleucine in position 9. These anchor amino acids are expected to be present in almost every K^d-bound peptide. One peptide was isolated from the K^d-restricted self-peptide pool and was determined by a combination of protein sequencing and mass spectrometry to be SYFPEITHI (Fig. 1). Besides the anchor amino acids, it carried several other motifs found in the pool like Pro^4 and Thr^7. Knowledge of these allele-specific binding motifs allows a forecast of possible CTL-epitopes: E. G. Pamer et al. [5] precisely predicted, according to the K^d-motif, the CTL-epitope of *Listeria monocytogenes*, GYKDGNEYI. Binding motifs have been determined for several MHC I types (Table 2); they can also serve as a basis for the construction of synthetic peptide sublibraries as discussed below.

Table 1: Pool-sequencing of a self-peptide mixture extracted from K^d-molecules [1]. The amount in pmol of 18 amino acids detected in each cycle is shown (Cys and Trp were omitted). Signals with twice the amount compared to the minimal amount in a previous cycle of the respective amino acid are considered significant and are underlined.

Pos	A Ala	R Arg	N Asn	D Asp	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	Y Tyr	V Val
1	173 1	0 3	15 6	9 0	15 2	19.6	241 6	1 8	43 2	42 1	94 8	15 4	37 7	22 0	172 0	57 1	40 6	84.6
2	37 7	1 5	4 5	4 2	6.7	8.9	128 8	1 4	10 2	12 6	10.1	3.8	34 7	8 2	24.2	10.1	291.4	14.8
3	45 8	7 5	24.5	7 8	15 0	27 1	140.5	2 3	80 2	128.1	27.9	9.9	19 7	6 4	14.8	10 3	33.8	43.6
4	69.3	3 0	18.5	13.2	27 1	15 8	103 8	5 0	10.4	15 0	12.5	4 0	1 9	103.6	14 6	11.2	6.9	17.0
5	49 9	1 5	23 2	17.4	10 4	13 4	98 5	1 9	18.7	27 6	3.8	23 9	1 2	36 2	31.8	19.3	3.5	40.9
6	45 9	9.7	14 2	10.0	8 4	10 7	67 2	4.0	26 6	28 0	58.6	26.4	5.2	10.1	6.5	7 6	8.4	32.2
7	11 3	1.2	38.3	12 4	20 3	22.9	25 4	3.0	4 0	7.1	9 6	3.5	1 2	4 3	15.2	53.5	2.8	12 7
8	14 3	3.9	13 3	11.3	27 6	16.9	25 3	5.9	2.8	10.7	9.5	2.3	5 7	3 2	12.4	14 3	13.3	25.8
9	4 9	2 3	6 5	6 2	8 2	3 3	10 5	1 7	38.3	34.5	0 5	1 7	1.5	2 3	3 2	3 6	3 1	9.9
10	2 8	2 2	2.7	4 3	3.7	1.9	6 8	1.1	13.2	17 7	0.4	1.1	1.5	1 8	1.8	1 5	1 7	3 7

Table 2: Binding motifs of two murine MHC I alleles, K^d and K^b [1]. Anchor amino acids are indicated in bold letters.

	The K^b-motif								The K^d-motif								
position	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9
anchor					**F** **Y**			**L**		**Y**							**I** **L**
strong signals			Y					M			I L N	P	M	K F	T N		
weak signals	R I L S A	N	P	R D E K T		T I E S	N K Q	I V	K A R S T V	F	V Q A M T E	A D E H N	N D S V T G	R H I M Y	E Q S	R E H F Y V	

Characterisation of the K^b-restricted synthetic peptide libraries

The optimized synthesis of mixtures representing a large variety of different peptides as well as the quality control of these peptide libraries and sublibraries require special and reliable analytical tools. We found that automated Edman degradation and electrospray mass spectrometry are well suited for this purpose. In order to evaluate the potential of these methods we selected one octapeptide sublibrary representing 48 possible MHC I-binding peptides. Due to our preliminary disappointing results in using more complex libraries in biological systems, we have chosen for this study a very small sublibrary. Based on the K^b-motif a set of 100 octapeptide mixtures was designed. Each mixture contained 48 analogs differing in their 6th to 8th positions. These peptides covered all possible combinations of amino acids occurring as so-called anchor residues and strong and weak signals from Edman degradation of a peptide pool isolated from the K^b-restricted cell line [1]. This synthetic "K^b-library" was prepared using p-hydroxymethyl-phenoxy-resin and standard Fmoc/tBu chemistry. To assure equimolarity of single components in the peptide mixture we used the divide, couple and recombine procedure [6]. Equimolar amounts of Fmoc-amino acid resins, loaded with Leu, Met, Ile or Val were mixed and Fmoc/deprotection was carried out with 50% piperidine in DMF for 15 min. After nine washing steps (3 x DMF, 3 x methanol, 3 x DMF) the resin was divided into three equal portions and coupling of Asn, Lys or Gln was performed in DMF using threefold excess of Fmoc-amino acids, TBTU, HOBt and sixfold excess of diisopropylethylamine (coupling concentration 0.35 M). The coupling reactions were complete (99 %) after 60 min as determined by Kaiser's ninhydrin test. The resins were filtered from the coupling reagents, washed with DMF and recombined for further washing and deprotection steps. This process was repeated for coupling of Thr,Ile,Glu,Ser and Phe,Tyr. The Y-, respectively F-(T,I,E,S)-(N,K,Q)-

(L,M,I,N) resins were divided into aliquots for coupling of defined amino acid positions 1-4 on a fully automated multiple peptide synthesizer (SMPS 350, Zinsser Analytic, Frankfurt, Germany) using tenfold excess of Fmoc-amino acids and the DIC/HOBt coupling procedure. The octapeptides were removed from the resin with trifluoroacetic acid containing 2.5% thioanisole and 2.5% thiocresol within 4 h at 25°C. The peptide solutions were filtered from the resins and peptide mixtures precipitated by the addition of cold ether and n-hexane. The precipitates were washed twice by sonication with n-hexane and were lyophilized from acetic acid/water/tert. butanol (1:10:50).

Analysis of a synthetic octapeptide sublibrary

Investigation by automated Edman degradation:

The synthetic mixture LNYRFX$_1$X$_2$X$_3$, containing only 48 octapeptides (T, I, E or S in position X$_1$, N, Q or K in position X$_2$ and L, M, I or V in position X$_3$) was subjected to automated Edman degradation. The results (Table 3) clearly proved the expected composition. Whereas the analysis of the first five positions revealed only one amino acid at a particular position (FNYRF), the expected amino acids at the X-positions could be unequivocally identified.

Investigation by electrospray mass spectrometry:

Electrospray [7] and pneumatically assisted electrospray [8] are soft ionization methods in which solute ions are emitted directly from charged droplets into the gas phase at atmospheric pressure. These techniques are especially suited for peptides, which form multiply charged quasi-molecular ions $(M+nH)^{n+}$. In the calculated mass ranges of the octapeptides of the library between 1012 and 1100 Da,

Table 3: Pool-sequencing of the synthetic octapeptide-sublibrary, consisting of LNYRFX$_1$X$_2$X$_3$ (X$_1$ = T,I,E or S, X$_2$ = N,Q or K, X$_3$ = L, M, I or V). Identified positions are underlined.

Pos.	A Ala	R Arg	N Asn	D Asp	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	Y Tyr	V Val
1	0.3	3.8	6.5	24.2	3.1	0.5	0.0	3.9	20.0	2860.2	4.5	2.3	6.8	1.8	10.4	280.5	4.8	5.4
2	0.0	0.0	4007.9	411.1	4.3	6.6	0 0	0.1	0.0	231.3	0.0	2.4	3.3	0.7	3.2	210.1	20.7	4 5
3	0.1	3.3	382.5	47.3	3.3	6.4	0 0	32.5	4 8	14.6	5.8	3.5	9.1	3.1	1.6	120.0	6709.4	5.2
4	0.3	1327.0	141.7	19.4	7.0	8.6	0 0	5.0	5.4	0.1	8.3	6.2	47.1	0.0	2.2	117.6	308.7	5.7
5	0.0	564.5	63.6	7.8	17.9	8.9	0.0	5.4	13 3	0.4	0.0	4 4	6749.1	0.1	5.5	81.0	30.9	6.2
6	0.7	152.9	37.0	18.5	1474.5	15.3	5.1	8.5	1019.1	2.6	12.8	4.9	132.3	0.0	532.3	1074.4	13.1	7.7
7	0.0	48.7	1571.3	151.5	463.6	1695.5	2 0	3 1	22 5	5.1	1546.5	20 2	20.8	0.3	13.8	10.1	6.8	12 7
8	0 0	47.1	92.5	11.2	31.8	64 1	0 6	5.4	456.2	416 8	45.2	695.3	9.1	1 7	2.7	55.6	4.1	973.0
9	0 0	13 1	12.3	2.8	6 3	6 6	0 0	0 8	0 0	0 2	0 0	17 8	36.2	0 7	0 9	46.4	2 3	19 6
10	0 1	22 9	10 1	0 0	5 5	6 0	0 0	0 8	3 4	3 2	1.5	0 0	0.0	0 1	1.0	39 2	0 2	4 9

are several possible amino acid combinations leading to an identical mass. In the present case, only 17 different mass values were obtained for the 48 peptides. As many as eight peptides (LNYRFLNT, LNYRFINT, LNYRFLKS, LNYRFIQS, LNYRFIKS, LNYRFVQT, LNYRFVKT and LNYRFLQS) of this mixture, for example, had a mass of 1040.2. The ion spray mass spectrum of this mixture recorded with a Sciex API-III triple-quadrupol mass spectrometer equipped with an ion spray interface (Sciex, Canada) showed the singly and doubly protonated molecular ions $(M+H)^+$ and $(M+2H)^{2+}$ of the individual peptides. The spectrum allowed an evaluation of the quality of the sublibrary. Firstly, because the relative intensities of an ion peak could be correlated with the number of peptides having this particular m/z value, the intensity of a peak (or better the peak area) that is correlated to the identical mass of eight peptides should be higher than that of a peak that is correlated to the mass of only four peptides (desirable but not always observed is a linear dependence between abundance and peak area). Secondly, masses which were found in the spectrum but deviated from the calculated masses of the expected peptides were caused by by-products formed during synthesis (failure sequences, incomplete removal of protecting groups, etc.). It should be emphasized that electrospray MS is only of limited use if the mixtures consist of too many (>1000) different peptides, because the different masses of the peptides can no longer be separated and distinguished from each other. In some of these cases, however,

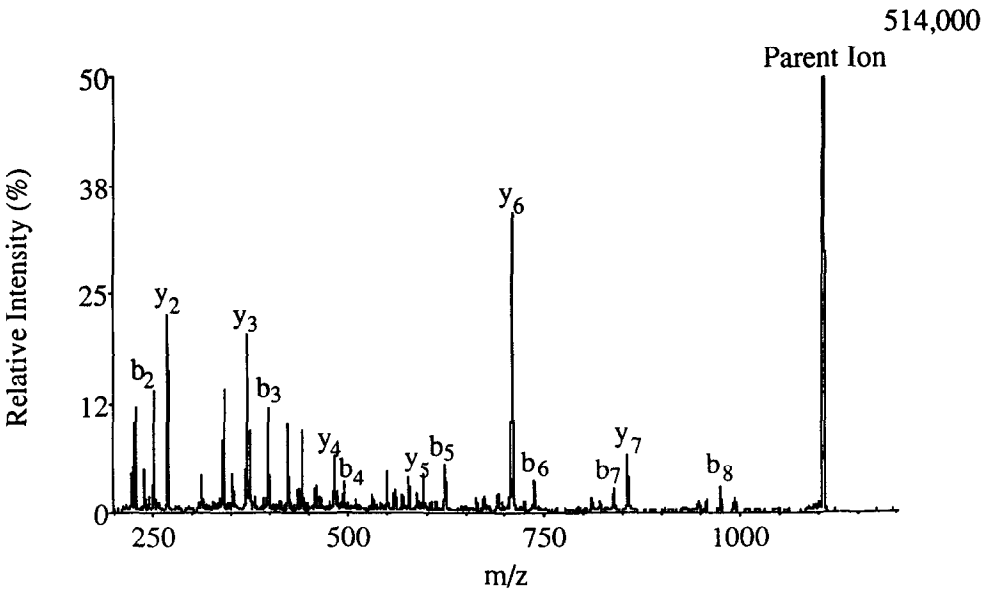

Fig. 1: Electrospray MS/MS of the epitope SYFPEITHI. The spectrum was obtained by collision induced dissociation of $(M+H)^+$ (M/Z 1106) with argon. Only the b and y ions are assigned (nomenclature according to Biemann [9]).

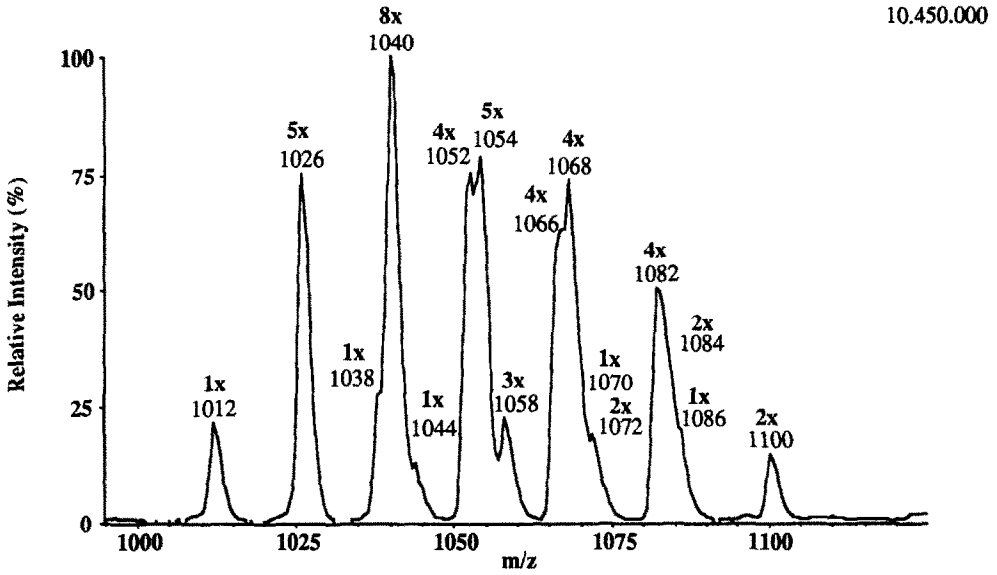

Fig. 2: Electrospray mass spectrum of a synthetic sublibrary. The spectrum shows the region of the singly charged quasi-molecular ions.

the efficiency could be enhanced by direct coupling of HPLC with the mass spectrometer (data not shown).

References

1. Falk, K.; Rötzschke, O.; Stevanovic, S.; Jung, G.; Rammensee, H.-G. Nature **1991**, 351, 290-296.
2. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. Nature **1991**, 354, 82-84.
3. Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. Nature **1991**, 354, 84-86.
4. Fodor, S. P.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. Science **1991**, 251, 767-773.
5. Pamer, E. G.; Harty, J. T.;. Bevan, J. M. Nature **1991**, 353, 852-855.
6. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. Int. J. Pept. Prot. Res. **1991**, 37, 487-493.
7. Whitehouse, C.M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. Anal. Chem. **1985**, 57, 675-679.
8. Bruins, A. P.; Covey, T. R. ; Henion, J. D. Anal. Chem. **1987**, 59, 2642-2646.
9. Biemann, K. Biomed. Environ. Mass Spectrom. **1989**, 16, 99.