

# **Search for Potential Vaccine Candidates by Identification of Tumor-Associated MHC class I Peptides using a High-throughput Mass Spectrometry Approach**

Caroline Ekblad  
SBC, Stockholm Bioinformatic Center  
19 September 2001

Arne Elofsson SBC, Stockholm Bioinformatic Center  
Rickard Glas, MTC The Karolinska Institute  
Ulrika Nilsson, Bo Karlberg Analytical Chemistry, Stockholm University  
Annika Persson, The Wenner-Gren Institute, Stockholm University

## **1 Introduction**

Immunotherapy of cancer has been of interest for long but there are still some problems to overcome before it can be used in routine preparative treatments. Better understanding of antigen processing and presentation, as well as recent advances in identification of new tumor antigens has paved the way for development of novel and specific vaccines including peptide-based vaccines. Of primary interest are tissue-specific gene products as well as these expressed or over-expressed in tumor cells but not in normal tissue. The aim of this project was to investigate whether it is possible to extract MHC class I specific peptides from tumor cells and in a high-throughput mode identify these peptides by use of mass spectrometry techniques. Peptides extracted from three different tumor-cell lines were analyzed. Candidates were also searched for by using the program SYFPEITHI, which have a function to predict MHC-binding motifs from a given sequence. The masses of the predicted peptides were calculated and compared to these obtained in mass spectra run on extracted peptides. Peptides with masses matching these predicted by SYFPEITHI were further analyzed by tandem mass spectrometry (MS/MS) for determination of their sequence. In order to validate the system, peptides with known mass and sequence were also analyzed. This approach, combining immunology knowledge, bioinformatics, and mass spectrometry to identify potentially MHC I (or MHC II) peptides can provide a powerful tool in the development of new vaccines. If optimized, it may allow rapid screening of a large number of peptides, including detection of low abundance peptides. This identification of peptides is based on the natural presentation of digested tumor-cell proteins without taking into account the immunogenicity of the peptides. Therefore, once a peptide is identified as a potential vaccine candidate, this peptide will be synthesized and tested in-vivo on mice. However, the advantage gained in this approach is the decreased number of peptides that need to be tested. An overview of the main steps to be performed in this approach is presented in figure 1.

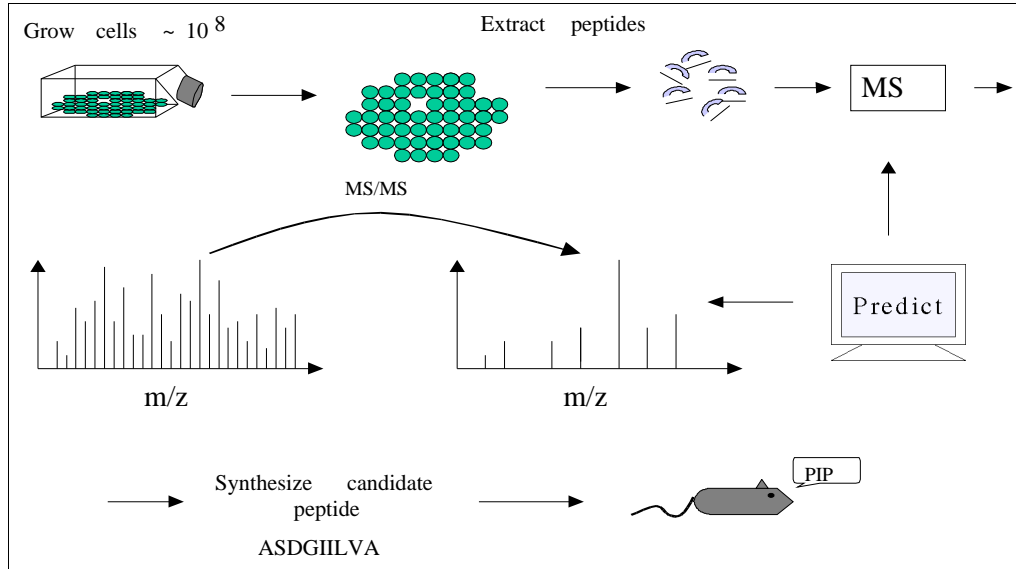


Figure 1. A schematic picture showing the planned procedure from growing of tumor-cells to injection on mice.

## 2 Background

The immune system has evolved a complex network of cells and organs in order to protect the individual against foreign invaders. It has the ability to distinguish between self and non-self and the effective specific immune system is based on the production of memory cells, which provide a fast response in case of a second infection. This allows vaccine treatments to be successful, since these induce a response that should be weak not to cause illness, but sufficient to generate memory cells. Traditional vaccines contain live, attenuated or whole, inactivated microbial organisms or antigenic subunits of these pathogens. The design of new vaccines has turned towards subunit or recombinant vaccines that contains only the components necessary to induce a specific immune response.

### 2.1 Antigen processing and presentation

Major histocompatibility complex (MHC) molecules are heterodimeric glycoproteins that are important in the cell-mediated immune response. They present a highly diverse set of peptides on the surface of a cell and induce T-cell activation. Proteins expressed in the cell are continually degraded by the proteasome complex to be reused in new synthesis. Peptides presented by the MHC class I molecules are mainly generated in the cytoplasm of the cell through the so called cytosolic pathway. Peptides that undergo complete degradation are transported by TAP proteins (Transporters Associated with antigen Processing) into the rough endoplasmic reticulum where 8- to 10- residue peptides are bound by MHC class I molecules. Binding of the peptide to MHC stabilize the molecule and enables transport of the complex to the cell surface (figure 2). This peptide driven mechanism prevents empty MHC molecules to be transported to the surface where they

otherwise might bind extracellular peptides. Non-self peptides bound to MHC class I proteins are recognized by cytotoxic T-cells leading to elimination of cells producing viral or mutant proteins [1, 2].

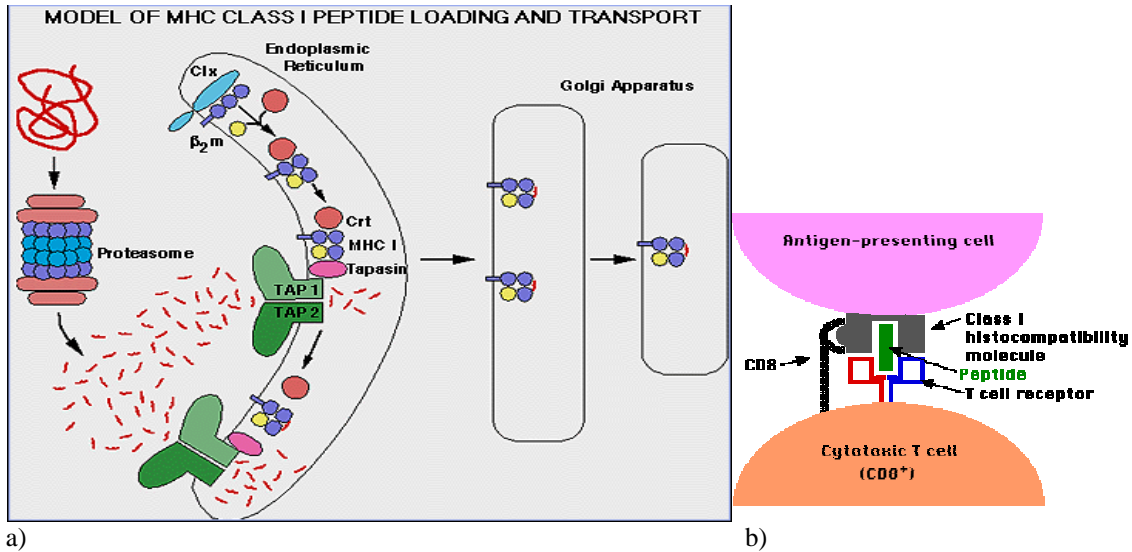


Figure 2. a) The protein is cleaved into short peptide fragments by the proteasome. The peptides are transported by the TAP complex into the endoplasmic reticulum where they bind to MHC class I molecules. The complex is transported via the Golgi apparatus to the cell surface where they are presented to T-cells. Image taken from Peter Cresswell, Yale University. [http://info.med.yale.edu/immuno/fac\\_cresswell.html](http://info.med.yale.edu/immuno/fac_cresswell.html)

b) A schematic picture showing the cells and molecules involved in the presentation of MHC class I peptides. Image taken from [http://www.ultranet.com/~jkimball/BiologyPages/H/HLA.html#Class\\_I\\_Histocompatibility\\_Molecules](http://www.ultranet.com/~jkimball/BiologyPages/H/HLA.html#Class_I_Histocompatibility_Molecules)

MHC class I molecules are found in almost all cell types present antigens to cytotoxic T-cells (CD8+) whereas MHC class II molecules are limited to antigen-presenting cells (APCs) present antigens to T-helper cells (CD4+). In addition to the CD4 and CD8 receptors (which bind the MHC molecule) the T-cells have receptors to detect foreign antigens. These T-cell receptors (TCRs) bind antigens only when they are presented by MHC molecules. Each T-cell express numerous but unique TCRs on its cell surface, and a TCR can only bind one specific antigen.

In humans, MHC class I gene contains three different loci, A, B and C, and the encoded proteins are designated HLA-A, HLA-B, and HLA-C (for Human Leukocyte Antigen). In mice, the corresponding loci are K and D and the encoded proteins are called H2K and H2D (for Histocompatibility-2). Further numbers or letters are used to specify the haplotypes of individuals [1].

## **2.2 Epitop prediction**

In order to design better therapeutic agents for infectious and autoimmune diseases it is helpful to predict which peptide can bind to MHC molecules and to understand their mode of binding. MHC class I associated peptides bind non-specific through their N- and C-terminal residues to the MHC groove. There are also specific interactions between side-chains and here certain anchor residues have been identified. These residues need not to be identical but related. They differ for peptides binding different MHC-types but are similar for all peptides binding the same MHC-type [1]. If the epitope, here referred to the antigen to which the MHC molecule binds and which are presented to T-cells, can be predicted, then the number of antigens likely to cause infection can be reduced. Computer algorithms for such predictions have been developed and these might help in identification of vaccine candidates. However, the programs up-to-date are limited by a number of factors including predications based on peptide libraries derived only from known protein antigens. This limits the number of MHC-types that can be significantly used for the prediction. Because of the difficulty to obtain sufficient experimental data there is a need to develop computational models of properties concerned with the interactions between the peptides and MHC. For instance, applying artificial neural network (ANN) based programs would reduce the number of laboratory experiments needed to elucidate the rules underlying the binding.

### **2.2.1 The SYFPEITHI database**

In this project the SYFPEITHI database freely available at <http://www.uni-tuebingen.de/uni/kxi/> was used. This database contains approximately 2000 peptide sequences reported for human, mouse, rat ape, cattle and chicken MHC class I and II types, taken from published data. As for MHC I epitope-prediction function, only alleles for which a large amount of data is available are included. This to increase the reliability which should be at least 80% in retrieving the most apt epitope. A naturally presented epitope is expected to appear among the top 2% high-scored peptides in more than 90% of the predictions. To-date the reliability for MHC class II peptides are lower (about 50 %) due to higher variability in the pocket-binding properties. The scoring system is based on which amino acids are present at each position, the most important being the anchor and auxiliary anchor positions [3].

## **2.3 Mass spectrometry**

During the past years much progress have been made in the development of new mass spectrometry techniques suitable for analysis of biomolecules. The main success can be described the development of novel ionization techniques which have made possible transfer of biomolecules from the liquid phase to the gas phase. Mass spectrometry is a powerful tool since determination of the molecular weight is made to a high degree of accuracy (~0.01-0.001%) and sensitivity (detection in the range  $10^{-9}$  -  $10^{-18}$  mol) [4].

A mass spectrometer consists of three major components: an ionization source, a mass analyzer and a detector. The mass analyzer separates the ions produced in the source region based on their mass-to-charge ratio (m/z). The ion detection system amplifies the ion current signal typically by an electron multiplier.

### 2.3.1 The Ionization Process

Compounds can only be analyzed in the gas phase, either as negatively- or positively-charged ions, which are generated by loss or gain of a charge. In peptide analysis this typically occurs via loss or gain of a proton ( $[M-H]^-$ ,  $[M+H]^+$ ). The ionization of the sample can be achieved in a number of ways. The most commonly used techniques in peptide analysis is Electrospray Ionization (ESI) and Matrix- Assisted Laser Desorption Ionization (MALDI). These are referred to as "soft" ionization techniques in the sense that the ionization process can take place without inducing fragmentation of the peptides.

In the case of **ESI** gas-phase ions are generated from molecules in solution. The sample is passed through a capillary tube and highly charged droplets are produced by applying a strong electric field to the end of the tube. As the sample flow through a slightly heated vacuum chamber solvent evaporates from the droplets causing an increase in the electrical charge density at the surface. At a critical point (the Rayleigh instability limit) the electrostatic repulsion overcomes the surface tension and forces a droplet to divide into smaller droplets. The final result is a singly or multiply protonated ions (depending on the size of the peptide as well as number of basic amino acids) which are accelerated by electric fields towards the mass analyzer [5].

In **MALDI** gas-phase ions are generated from a solid phase. The principle of MALDI is based on all molecules naturally possessing rotational, vibrational and electronic energy (if in a liquid or in a gas even kinetic energy). If energy is supplied in a large amount over a short period of time the energy can not be dissipated to the surrounding fast enough. Therefore, this increase of internal energy causes vibrational and rotational energy to turn into translational energy leading to melting, vaporization and some ionization. The energy source in MALDI is a laser beam which by focusing can deposit a large density of energy into a small space. However, for a substance to absorb a specific energy it must have a matching absorption spectra [6]. Therefore, the compound to be analyzed is mixed and co-crystallized with a photoactive matrix that absorbs light at the laser wavelength used. Matrix should be present in excess in order to isolate individual analyte molecules. The exact mechanism behind protonation of the analyte molecule has been debated, but are most likely formed in the gas phase by transfer of protons from protonised matrix molecules to neutral analyte species. The matrix is protonated through photo-ionisation of highly excited matrix molecules forming radical ions. When these interacts with neutral matrix molecules protonated matrix molecules and neutral matrix radicals are formed [5]. Multiply charged analyte ions are prevented because of proton-transfer reactions with reactive matrix molecules and capture of free electrons. For molecules less than 10,000 Da, a-cyano-4-hydroxycinnamic acid is commonly used as matrix material.

One major advantage of ESI over MALDI is that a liquid-chromatograph can be directly coupled to the ESI interphase enabling on-line separation of the sample (LC-MS). Besides that peptides are being separated they also become concentrated, thus improving the level of sensitivity. In ESI multiply charged ions are generated which make possible accurate mass measurements. However, this increases the complexity of the spectrum and complicates the interpretation, especially if a mix of many different peptides is being analyzed. In this case MALDI, which mainly generates singly charged ions, might be

preferable. MALDI is also less sensitive for salts.

### 2.3.2 The Mass Analyzer

Various combinations of ionization sources with mass analyzers are possible. However, ESI is commonly coupled to an ion-trap or a quadrupole time-of-flight (TOF) analyzer and MALDI is often used in combination with a TOF analyzer. Each mass analyzer has its own special characteristics and here will be given only a brief overview of the different analyzers used in this project.

The **ion-trap analyzer** works by isolating selected ions in a confined region of space which is then manipulated by using a radio frequency electric field (RF). Ions are sorted by changing the field conditions to destabilize ions of a particular mass which will then leave the trap and be collected by the detector.

In a **quadrupole** instrument, electric fields are used to separate ions as they pass along the central axis of four parallel, equidistant rods which have static (DC) and radio frequency (RF) voltages applied to them. The strength and the frequency of the electric field is changed so that only ions of selected mass/masses can be transmitted, others are deflected to strike the rods.

A **time-of-flight** analyzer separates ions given the equivalent kinetic energy on the basis of mass. Lighter ions will have a higher velocity than heavier ones and will therefore reach the detector sooner.

### 2.3.3 Peptide Fragmentation

Tandem mass spectrometry (MS/MS) can be used to gain information about the peptide sequence. A specific molecular mass is selected and subjected to collision-induced-dissociation (CID) generated by bombardment with an inert gas. What fragments ions that will appear is dependent on factors as primary sequence, charge state, the amount of internal energy, how the energy was introduced etc. For a fragment to be detected it needs to carry at least one charge and the resulting ions are classified according to where the charge is retained. If it appears on the N-terminal fragment fragments are designated as either a, b or c ions and if it appears on the C-terminal fragment as either x, y or z ions depending on the bond cleaved (figure 3). Also internal ions, resulting from double backbone cleavage, can appear. In addition to the formation of the above ions, losses of H<sub>2</sub>O and NH<sub>3</sub> are very common. In low energy CID a, b and y ions are the most abundant fragment ions, whereas in high energy CID all of the above ions are generated [5]. On a MALDI-TOF system fragmentation is achieved by post-source-decay (PSD) and optionally in combination with CID. Since the rules for fragmentation of peptides are quite complex there is a need for computer programs to help to interpret MS/MS spectra.

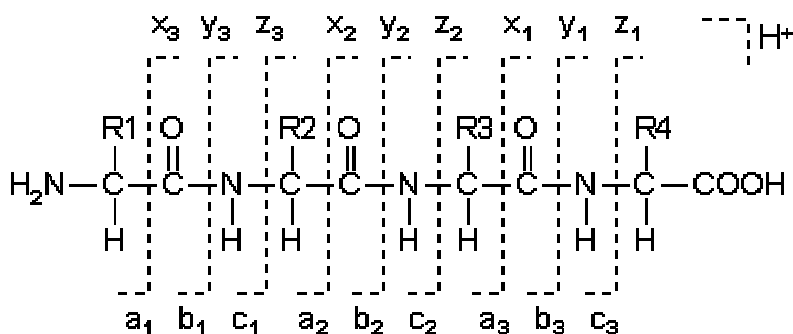


Figure 3. Ion series nomenclature first proposed by Roepstorff and Fohlman (1984) and later modified by Biemann (1988) [5].

### 2.3.4 LC-MC

The analysis of complex sample solutions is preferentially aided to some separation before applied to the mass spectrometer. This can be performed by using reversed-phase high performance liquid chromatograph (HPLC) which separate peptides according to hydrophobicity. This method is well suited since it can be interfaced to an ESI ion source and thereby allow online separation (commonly referred to as LC/MS). Separations in reversed-phase chromatography is based on reversible adsorption of solute molecules with varying degrees of hydrophobicity to a hydrophobic stationary phase. Elution is achieved by increasing the polarity of the mobile phase by adding an organic solvents, as for instance acetonitrile. In order improve peak sharpness, ion-pairing agents as trifluoroacetic acid (TFA), formic acid, or acetic acid are often included. The effect is likely due to the ability of the ion-pairing agent to mask positive charged and polar groups on proteins and peptides and thereby enhance interaction with the hydrophobic reversed-phase surface [7]. Columns used in reversed-phase chromatography are commonly packed with silica-matrix beads with coupled hydrophobic ligands.

## 3 Methods

### 3.1 Growing and harvesting of cells

Three different cell lines were grown: HeLa cells (isolated in 1951 from the patient Henrietta Lacks) containing multiple copies of the human papillomavirus type 18 (HPV-18) responsible for causing cervical cancer; Caov4 a human ovarian-cancer cell line; and B16-B7 melanoma mouse cells visibly producing melanin causing skin cancer. The two human samples are HLA-A2 specific and the mouse sample H2-Kb specific. Cells were grown until a density of  $\sim 10^8$  cells. Cell medium was removed. 1x Trypsin-EDTA was added (covering the bottom) and the flask placed in 36.6°C incubator until colonies float off. The flask was rinsed with PBS. The solution was transferred to a 50-ml falcon tube and centrifuged for 5 min at 2000 g. The pellet was resuspended in PBS. Cells were counted and centrifugated for 5 min at 2000 g. The pellet was frozen (-80°C) or used directly for peptide extraction.

### **3.2 Extraction of peptides**

Two different methods were applied to extract the peptides. First a "cell-stripping" method was applied to the HPV18 sample. In this method peptides are simply eluted from the surface of the cells. Later all three samples were subjected to immunoprecipitation in which the peptides react specifically with an antibody and are removed from the solution.

#### **3.2.1 Cell-stripping**

HeLa cells were washed three times in 25 ml PBS. Stripping was performed by addition of 20 ml PBS adjusted to pH 2 by TFA and incubation for 2 min at room temperature. The sample was centrifuged and the supernatant saved.

#### **3.2.2 Immunoprecipitation**

Cells were lysed in 1 ml NP-40 buffer (10mM tris buffer pH 7.8, 150 mM NaCl, 5mM MgCl, 1mM EDTA, 0,5% NP40, 0.1% protease inhibitor cocktail) rotating for 30 min at 4°C. The sample was spun for 10 min at 14,000 rpm. To reduce background the lysate was precleared by adding of 5 ml NMS (Normal Mouse Serum) and 50 ml protein A sepharose beads per ml of lysate, followed by incubation for 60 min rotating at 4°C. The sample was spun for 10 min at 14,000 rpm. To the supernatant 20 ml BB7.2 (HLA-A2 specific) or a-8 (H2-Kb specific) antibodies were added to the human and mouse samples respectively. Incubation for 60 min rotating at 4°C was followed by addition of 50 ml protein A sepharose beads and incubation for additionally 45 min. The samples were spun for 4 min at 14,000 rpm and the pellet washed twice in cold NET-buffer (150 mM NaCl, 50 mM Tris pH 7.5, 5 mM EDTA, 0.5 % NP-40). The pellets were solved in 200 ml dH2O with 0.1 % TFA. The samples were spun for 10 min at 14,000 rpm. The HPV18 sample was subjected to three additional washing steps in order to get rid of salt that interfere with the ESI technique. The supernatant was filtered in MICRON 3K filters.

### **3.3 Prediction of MHC-binding peptides**

Prediction of potential HLA-2 and H2-Kb peptides binding to MHC class I molecules was carried out by using the SYFPEITHI database (Ver 1.0) freely available at <http://www.uni-tuebingen.de/uni/kxi/>. Input sequences were down loaded from NCBI (<http://www2.ncbi.nlm.nih.gov/>). As for HPV18, for which the complete genome is known, all the encoding protein sequences (64) could be down loaded from <http://www.ncbi.nlm.nih.gov:80/PMGifs/Genomes/vis.html>). Regarding the other two samples manual searches were performed in the NCBI protein database. As for the human Caov4 strain the search "ovary AND homo sapiens" generated 806 sequences, "ovarian AND cancer AND homo sapiens" 235 sequences and "ovarian cancer AND homo sapiens" 174 sequences. In order not to loose important data the 806 sequences were down loaded and used as input for the SYFPEITHI epitope prediction. In the case of the mouse melanoma cells 154 sequences were down loaded, generated from the search "melanoma AND mouse".

For simple input and extracting of relevant data a perl program, *epitoppred.pl*, was constructed (appendix 1). It takes any file in FASTA format, extracts an ID of each protein and passes on each sequence to the SYFPEITHI epitope predictor. From the result

file all peptides with a scoring equal to or higher than 20 are selected. For these peptides the monoisotopic, the  $[M+2H]^{2+}$  and the  $[M+3H]^{3+}$  mass is calculated as well as the pI. The pI was calculated using the pI/Mw tool provided by the ExPASy Molecular Biology server of the Swiss Institute of Bioinformatics found at <http://www.expasy.ch/tools/>. Along with the pI it also calculates the average mass of a peptide or protein and to make possible comparison with the monoisotopic mass this average mass is also listed in the output result file. The generated output file contains the protein ID, the start position of the peptide in the protein sequence, the MHC allele, the peptide sequence, the score and the masses and pI as described above. Two different output files are created: one sorted according to scoring and one sorted according to mass.

For the human samples all HLA-A2 MHC alleles available in the data base were selected (HLA-A\* 0201, HLA-A\*0202, HLA-A\* 0203) and prediction performed both with nonamers and decamers. For the mouse sample the H2-Kb allele was selected and predictions performed with octamers.

### **3.4 Mass spectrometry performance**

Three different instruments, an ESI-LC-Q-ion trap, a MALDI-TOF and an ESI-Q-TOF, were used to collect MS and MS/MS spectra of the extracted peptide mixtures and single peptides with known sequences.

#### **3.4.1 LC-MS**

Peptides were analyzed with a reversed-phase HPLC system connected to an ESI-equipped quadrupole ion trap mass spectrometer (Thermo-Finnigan). A XTerra™ MS C-18 column, bead size 3.5 mm and with the dimensions 1.0 x 100 mm was mainly used and will be referred to as the C-18 column. (Initially a column with dimensions 2.0 x 150 mm was tried, and this will be referred to as C-18\*) 2 or 5 ml sample were loaded on to column through an injection loop. Different solvents were tried as well as various gradients and flow rates:

- 1) Solvent A: 4mM NH<sub>4</sub>Ac, adjusted to pH 3 with formic acid  
Solvent B: 2mM NH<sub>4</sub>Ac, 70% AcN, adjusted to pH 3 with formic acid
- 2) Solvent A: 100% dH<sub>2</sub>O, 0.1% TFA  
Solvent B: 100% AcN, 0.1% TFA
- 3) Solvent A: 100% dH<sub>2</sub>O, adjusted to pH 3 with acetic acid  
Solvent B: 100% AcN, adjusted to pH 3 acetic acid

Usually linear gradients from 0 to 100% B within 1-100 min were applied. Flow rates were varied between 2 and 500 /min not exceeding the maximum operating pressure of 400 bar. Tuning was made with the following settings (previous optimized for a known octapeptide): capillary temp: 185°C, capillary voltage: 10 V, tube lens offset: 15 V, spray voltage 5 kV and ions were analyzed in positive mode. The peptide concentrations of the mixed samples were unknown but different dilution series were tried. For known peptides a concentration between 1pmol/ml and 1nmol/ml was analyzed.

In MS/MS analysis collision energies between 15-35 were applied, depending on the ease

in fragment the peptide.

### **3.4.2 MALDI-TOF**

Samples were prepared in 70:30 % AcN: dH<sub>2</sub>O, 1% TFA and mixed 1:1 (0.5 +0.5 ml) directly on a 100 well sample plate with saturated a-cyano-4-hydroxycinnamic acid matrix prepared in 70:30% AcN: dH<sub>2</sub>O, 1% TFA. The sample plate was allowed to dry for about 20 min. Spectra were collected on a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems). Standard instrument settings for angiotensin provided with the system were loaded and changed if needed. In general ions were analyzed in positive reflector mode with a grid voltage set to 70%. Laser intensity was adjusted as appropriate for each specific sample and spot. Calibration was performed internally or externally by use of a standard calibration mix. Usually 300 shots at 3-6 positions were collected and summed.

PSD-analysis was performed by acquisition of different segments over the mass range for which fragments could be observed. The files were then stiched together to generate a result spectrum.

### **3.4.3 Q-TOF**

A few experiments were performed on a ESI-equipped Q-TOF instrument (MicroMass) in collaboration with Annika Persson, WGI. Samples were prepared in 50:50% AcN: dH<sub>2</sub>O, 0.1% / 1% HAc or 0.1% TFA. Both MS and MS/MS analysis were performed and instrument settings optimized as needed. Unfortunately no online separation has yet been set up for this instrument. Samples were loaded directly in a needle later assembled in the ESI interphase.

## **4 Results**

As expected the eluted tumor-cell samples, assumed to contain 1,000-10,000 peptides, gave very complex spectra. If the mixture is not prefractionated or separated online far to many peptides will appear in the spectrum and MS/MS analysis will be almost impossible to perform since too many ions will be fragmented at the same time. For instance, in MS/MS mode the parent ion mass is given but in order to account for instrument errors ions within approximately  $\pm 0.4$  Da will be collected and then fragmented. The main problem was to separate the peptides with good resolution using the equipment available. Since it turned out to be hard to get any relevant data out of the complex mixtures, the attention turned to analyze known peptides and mixtures thereof in order to find a LC/MS system that works reliably.

### **4.1 LC-Q MS**

As mentioned above, much effort was put into finding a setup that could separate the peptides with acceptable resolution. For the C-18 column used a decreasing in flow rate below ~10 ml/min did not improve resolution since it gave taling peaks. Flow rates above 100 ml/min often gave leakage in the system. Best separation conditions were found when run over a long gradient as 60 min.

Of much help would be if the elution time for a certain peptide could be properly calculated. This would help tracing high-scored peptides predicted by the epitop-predictor program. Since peptide retention is dependent on the amino acid composition such calculations can be done by summing the relative hydrophobic contributions from each amino acid. Also peptide length have proven to be important making the prediction more complex []. Two different programs predicting elution pattern according to hydrophobicity were investigated: AminoXpress (freely downloaded from [http://www.legend2000.com/aminopress/ami\\_index.asp](http://www.legend2000.com/aminopress/ami_index.asp)) and GPMAW Ver 4.11 (a commercial program from Lighthouse data). However, neither of the two were able to predict an elution profile observed for known peptides. Both programs are very simple and desirably more parameters should be given as input, as for instance column properties.

#### **4.2 MALDI-TOF**

Data collected on tumor-derived peptides did not give any good spectra. HPV18 only gave a noisy signal and peptides from Caov4 and B16-B7 gave a pattern likely caused by adducts. When singly or mixed known peptides were analyzed well defined peaks were observed, though some identified with salt bound.

#### **4.3 Q-TOF**

Best quality data was obtained on the Q-TOF mass spectrometer due to the high sensitivity of the instrument. However, because lack of time only a few analysis were made and HPV18 was the only tumor-derived samples analyzed. As expected, this generated a very complex spectra. Even if each peptide can be seen the resolution is not good enough to evaluate charge state and MS/MS analysis of a selected peak is not likely to generate a reasonable spectrum.

When a mix of 10 known peptides were analyzed peptides with multiple Arg, Lys or His were not seen as singly charged ions.

#### **4.4 MS/MS analysis**

Fragmentation were induced on all three instruments used. A problem observed on the ESI-Q iontrap mass spectrometer was how to find an appropriate collision energy to obtain fragments over a wide mass range. Desirably 10% of the parent ion should be left. Very often only one or two amino acids were cleaved off generating a single dominant peak and only minor peaks of smaller fragments. Such data are not convenient for automatic interpretation of spectra. When this was observed fragmentation of the daughter ion was also tried. On the instrument used MS<sup>n</sup> experiments can be performed up to n=9. The problem might also be overcome if instead the corresponding doubly charged ion is subjected to fragmentation. Then additional charged fragments will form. However, this requires some deconvolution of the spectrum to only singly charged ions.

PSD analysis on the MALDI-TOF was tedious and for some segments no data could be collected which spoils automatic interpretation of spectra. Only known peptides were analyzed and MassLynx PepSeq program was used for interpretation. Only a few amino acids could be correctly identified.

On the Q-TOF, one MS/MS sequence analysis was successfully performed using MassLynx PepSeq program. The sequence was not fully automatically predicted, but needed a human eye as well. Also, the peptide was not one of the known, but a failed synthesis product thereof.

Another program for de novo interpretation of peptide CID spectra is Lutefisk, developed by Richard S Johnson, Immunex Corporation (1997-2000). This software is freely available at <http://www.immunex.com/researcher/lutefisk/>. Currently, this program is limited to interpret data from low energy CID of tryptic peptides on a triple quadrupole, a Q-TOF and an ion trap. However, there is still an option to select no digest for the proteolysis parameter. When data from LC-Q MS/MS spectra were imported the reply was almost always "don't waste your time on this one". However this might well be due to bad input data.

MassLynx is much more powerful and has an interactive userphase. Data can be imported in text file format but then automatic prediction can not be applied.

#### 4.5 Spectra

A few spectra have been selected to illustrate the obtained results. These are presented in appendix and comments will be given below. Other run spectra are collected in an order given to Arne Elofsson, SBC as is a CD with all raw data files saved.

##### The known peptides analyzed:

Peptide	Sequence	Mass (Da)
E11	CIDGVCWTV	995.18
E16	ILDSFDPLV	1018.17
E14	SLMAFTAAV	910.10
KAP1	MDRGLTVFVAVHVPD	1655.93
KAP3	VHVPDVLLNGWRWRL	1860.19
KAP10	RGPVAFRTRVATGAH	1595.83
ORFGLA	GLAAATWVWL	1087.29
ORFYQL	YQLPVVFGV	1035.25
ORFLLM	LLMSVVVAGL	1001.29
ORFFVF	FVIFYQLFVV	1161.41

#### 1) Illustration of the complex peptide mixtures derived from tumor-cell lines

##### a) HPV18 (HeLa cells) LCQ iontrap

Solvent A: 100% dH<sub>2</sub>O, 0.1% TFA

Solvent B: 100% AcN, 0.1% TFA

Gradient 0-100% B in 60 min, flowrate 2ml/min, C-18

##### b) HPV18 Q-TOF same sample as in a).

##### c) Caov4

100% Solvent A: 4mM NH<sub>4</sub>Ac, adjusted to pH 3 with formic acid

Flow rate 50 ml/min. C-18\*

**d) B16-B7**

Same conditions as in c)

Where separation on a column was used peaks are summed over the whole elution range. No separation of peptides was performed in b). Each peptide is resolved, but charge state is not easily evaluated. Comparison between a) and b) agree in the zoomed range 860-870 (mass and peak heights). At this resolution it is however difficult to tell if the peaks observed in c) and b) are other peptides or the same just dependent of the instrument settings. Some characteristic peaks are observed in both c) and d): 673.2, 695.1, 717.1, 739.2. These might origin from non cell-specific proteins. Most peptides presented by MHC class I molecules are expected to be in the range 800-1200 Da, but can vary from 600-1500.

**2) LC-Q on a mix of the 10 known peptides above**

Solvent A: 100% dH<sub>2</sub>O, adjusted to pH 3 with acetic acid

Solvent B: 100% AcN, adjusted to pH 3 acetic acid

0-100% B in 30 min, flow rate 100 ml/min

Peptide concentration 1 nmol/ml (each peptide)

All peptides but ORFYQL could be identified. The fail in identify this peptide is probably only because of miss calculation of the mass and the peak is likely to be the one between 13.64 and 14.45 min. When same sample is run over a 60-min gradient the peaks are further resolved but the peak sharpness reduced. The latter might be overcome using 0.1% TFA instead of acetic acid.

**3) MALDI on a mix of the 10 known peptides above**

Peptide concentration 10pmol/ml (each peptide)

The spectrum illustrates the variability in ionization properties of different peptides and encourage carefully interpretation. Spectra of each individual peptide was collected as well. Most peptides could be identified with or without modifications.

**4) Q-TOF KAP1**

Sample in AcN:H<sub>2</sub>O 50:50% 0.1% HAc. Sample concentration 1pmol/ml

The original spectra was poor, but a doubly charged ion was found at 643.35 Da. It did not agree with the expected parent ion but was subjected to CID analysis anyway. The sequence solved is MCRFVAVHVPD, which is KAP1 with some missing amino acids within the peptide (failed synthesis). The parent mass for this peptide is 1286 and this peak was identified in MALDI spectra as well.

The few peaks seen in the initial spectra is due to low concentration. When a sample at concentration 10 pmol/ml was analyzed more peaks appeared. However, the parent ion at 1655.93 Da was not seen. The explanation is probably found in the sequence of the peptide. If Arg and or His appear in their charged state no singly charged ions will be

seen.

## **5 Discussion and future prospects**

Since MALDI only generates singly charge ions, such a spectrum in general becomes less complex. Therefore, it might be worth to collect some initial spectra on an instrument with this interphase. Also, the difference in ionization techniques between ESI and MALDI render possible that ions not observed in ESI might be observed in MALDI and vice versa.

The ease of interpretation of MS/MS data is to some degree dependent on what instrument is used. The most convenient is probably an ESI-Q-TOF since it produces very high-resolution and high-mass-accuracy full-range spectra of the daughter ion. The charged states of the product ions are easily distinguished and deconvolution of the spectrum to one containing only singly charged ions render impossible interpretation of higher order charge. Furthermore, the mass accuracy allow to distinguish between Gln and Lys differing only 0.03 Da in mass. The iontrap instrument is limited in the number of ions it can trap between the quadropols without losing resolution and mass accuracy. One advantage with the trap though is the possibility to retain the daughter ions after MS/MS analysis and fragmentation can proceed ( $MS^n$ ). Fragmentation of highly charged ions might cause difficulties when interpreting the spectrum since internal fragments are likely to be seen. In this case also a less informative spectrum can be obtained since presence of charged amino acids in the middle of the peptide may favor cleavage at specific bonds. When PSD analysis is concerned it can be used to confirm a peptide sequence but is not the method of choice when analyzing unknown peptides. Analysis pattern of a given spot as well as the laser intensity for each segment to be collected is easily programmed. However, since the energy needed to induce fragmentation is sequence dependent it might be difficult to set an appropriate laser intensity for each segment, especially since it is not always increasing in a linear manner. The quality of the data obtained from induced fragmentation is also dependent on the peptide for instance how easy bonds are broken and the presence of charged amino acids.

Manual interpretation of MS/MS data is very hard and tedious task. Even when automatic sequence programs are used one should expect to spend hours to solve a sequence. The programs often suggest more than one possible sequence and a close examination of peaks is needed. Especially this applies when no enzymatic cleavage has been performed. This requires a wide knowledge of the theories behind fragmentation as well as awareness of all chemical modifications that might have taken place. The ultimate proof of a sequence assignment is to synthesize the peptide and compare the spectra.

There is probably also a need for optimizing the peptide-extraction procedure. As less salt as possible should be present in the final sample since this suppress ionisation in ESI.

Better resolution of HPLC separation would possibly be obtained with a longer column, and more important, with smaller beads.

As for instrumentation, the method of choice is a Q-TOF with on-line separation. It might be worth trying off-line separation and analyze fraction by fraction if this desired instrument set-up is not available. To further increase resolution in separation of peptides, capillary electrophoresis is recommended.

Another interesting approach is to start from the predicted epitopes. If a high-scored peptide is interesting from a medical point of view an idea is to synthesize this peptide. Next step is then to identify this in the tumor-derived sample which is simplified since elution time can be predicted. Also fragmentation pattern of the two can be compared.

A general problem in identification of a vaccine candidate peptide is that peptides derived from pathogens are often present in much lower concentration than naturally processed self-peptides. Therefore this signal might be "suppressed" in any spectra and hard to detect.

Interesting would also be to analyze sample of both malignant and non-malignant tissue from the same patient and search for distinctions. The problem though is to get access to such samples as well as get the normal tissue cells to grow sufficiently.

Once a peptide is identified the final work is to establish the immune response actually caused by this peptide. If it appears to be an evil one the way to immunotherapy of cancer is somewhat shortened.

1. Charles Jr. Janeway, *Immunobiology : The Immune System in Health and Disease* Blackwell Science Inc; ISBN: 0865428115
2. Kenneth L. Rock and Alfred L. Goldberg, *Degradation of cell proteins and the generation of MHC class I-presented peptides*, *Annu. Rev. Immunol.* 1999. 17:739-79
3. Hans-Georg Rammensee, Jutta Bachmann, Niels Nikolaus Emmerich, Oskar Alexander Bachor, Stefan Stevanovic: *SYFPEITHI: database for MHC ligands and peptide motifs*. *Immunogenetics* (1999) 50: 213-219 (access via : <http://www.uni-tuebingen.de/uni/kxi/>)
4. Poland GA, Ovsyannikova IG, Johnson KL, Naylor S. *The role of mass spectrometry in vaccine development*. *Vaccine*. 2001 Mar 21;19(17-19):2692-700
5. P. James (Editor) *Proteome Research : Mass Spectrometry (Principles and Practice)* Springer Verlag; ISBN: 3540672567
6. <http://www.micromass.co.uk/basics/> (17 September 2001)
7. Vydac Advances Newsletter for Spring, 1997, *What is 0.1% TFA* <http://www.vydac.com/vydacpubs/NL97SPR/NL97SPR.html>
8. Chabanet C, Yvon M. *Prediction of peptide retention time in reversed-phase high-performance liquid chromatography*. *J Chromatogr.* 1992 May 22;599(1-2):211-25.